

SEQ ID NO: 18
Database: N-Geneseq-032802
Ac. NO: AAV99731

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2002, 02:17:00 ; Search time 272.98 Seconds
(without alignments)
3025.261 Million cell updates/sec

Title: US-09-746-783-18

Perfect score: 481
Sequence: 1 GGATGCTGTAATAATAGCA.....AAAAAAAAAAAAAAAAAAAA 481

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/genesqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	481	100.0	481	20	AAV99731
2	289.6	60.2	318	20	AAV88049
3	70.4	14.6	631	21	AAI16169
4	61.6	12.8	382	23	AAI23536
5	61.6	12.8	516	21	AAI03962
6	60.6	12.6	581	10	AAI93083
7	60.6	12.6	581	20	AAI20544
8	60	12.5	557	21	AAI69778
9	60	12.5	594	21	AAI69777

ALIGNMENTS

10	60	12.5	601	21	AAI69823	Human ovarian carc
11	60	12.5	624	21	AAI16521	Human colon cancer
12	60	12.5	630	14	AAI35289	Human TC-CGF. Hom
13	60	12.5	630	22	AAI09895	Human T cell -deriv
14	58.4	12.1	456	14	AAI49402	14kDa BCEF/thiore
15	56	11.6	315	14	AAI43433	ADP gene. Homo sa
16	56	11.6	315	15	AAI087818	Sequence encoding
17	54.4	11.3	315	15	AAI051112	Human ADF-polypept
18	54.2	11.3	605	21	AAI16480	Human colon cancer
19	53.6	11.1	332	15	AAI49403	reCFE 104 coding s
20	53	11.0	312	15	AAI05111	Human ADF-polypept
21	53	11.0	312	16	AAI087817	Sequence encoding
22	52.6	10.9	260	14	AAI49405	reCFE 84 coding se
23	52.6	10.9	272	14	AAI49404	reCFE 84 coding se
24	50	10.4	562	19	AAI64661	Seq ID 29 from DEL
25	49.2	10.2	681	24	AAI62239	CDNA sequence #26
26	49.2	10.2	5675	24	AAI33276	Human immune syste
27	49.2	10.2	11694	22	AAI6697	Tumour suppressor
28	48.8	10.1	1623	17	AAI14925	T cell replacing f
29	48.8	10.1	1623	19	AAI64062	Plasmid pSP6K-mRFP
30	48.8	10.1	10369	22	AAI546303	Tumour suppressor
31	48.8	10.1	10369	22	AAI32392	Human immune syste
32	47.4	9.9	345	22	AAI70126	Human cervical can
33	47.4	9.9	1038	20	AAI84970	Human secreted pro
34	47.4	9.9	1203	20	AAI24424	Human bladder tumo
35	47.4	9.9	1463	22	AAI08459	Human secreted pro
36	47.2	9.8	278	22	AAI10134	Human breast can
37	47.2	9.8	2771	22	AAI72860	Human cervical can
38	47	9.8	695	22	AAI1586	Human breast can
39	47	9.8	781	22	AAI24400	Human breast can
40	47	9.8	5173	18	AAI89783	Human immunodefici
41	46.8	9.7	555	22	AAI27239	CDNA encoding nove
42	46.8	9.7	1701	21	AAI33336	Human secreted pro
43	46.8	9.7	2406	21	AAI69953	Human secreted pro
44	46.8	9.7	1897	24	AAI32570	Human immune syste
45	46.6	9.7	462	22	AAI84644	Human polynucleoti

RESULT 1

AAV99731 standard; CDNA; 481 BP.

AC	AAV99731:		
XX			
DT	26-APR-1999	(first entry)	
XX			
DE	Human adult testis secreted protein fg505_4 cDNA.		
XX			
KW	Secreted protein; human; testis; fg505_4; ds.		
XX			
OS	Homo sapiens.		
XX			
FT	Key	Location/Qualifiers	
XX	CDS	84..407	
XX		/*tag= a	
XX			
XX	WO9856909-A2.		
XX			
PD	17-DEC-1998.		
XX			
PF	08-JUN-1998;	98WO-US11822.	
XX			
PR	05-JUN-1998;	98US-0092722.	
XX	11-JUN-1997;	97US-0873218.	
PA	(GENM) GENETICS INST INC.		
XX			
PI	Agostino MJ, Fechtel K, Howes SH, Jacobs K, Lavallie ER,		
XX	McCoy JM, Racie LA, Spaulding V, Treacy M;		

DR MPI: 1999-080899/07.
DR P-PSDB; AAM95351.

PT New polynucleotides encoding secreted human proteins - derived from
PT human foetal brain, adult testes, foetal kidney, adult thyroid or
PT adult retina cDNA libraries

PS Claim 30; Page 85; 113pp; English.

CC This is the nucleotide sequence of cDNA clone fq505.4, which
CC includes an open reading frame for a 107-amino acid polypeptide
CC (see AAM95351). The clone was isolated from a human adult testis
CC cDNA library using methods which are selective for cDNAs encoding
CC secreted proteins, or was identified as encoding a secreted or
CC transmembrane protein on the basis of computer analysis of the
CC amino acid sequence of the encoding protein. Database searches
CC indicate some sequence similarity to known sequences. The
CC invention provides cDNA clones (see AAV9721-33) from human adult
CC thyroid, adult retina, adult testis, foetal kidney and foetal brain
CC that encode novel secreted proteins (see AAM95344-53). Each clone is
CC individually available from deposit clone ATCC 98451 (see also
CC AAV9734-43). The isolated polynucleotides (PNS) and proteins are
CC predicted to have activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans
CC and animals, although no supporting data is given. Suggested
CC activities include nutritional, cytokine, cell proliferation or
CC differentiation, immune stimulating (e.g. as vaccines) or immune
CC suppressing, haematopoiesis regulating, tissue growth,
CC activin/inhibin, chemotactic/chemokinetic, haemostatic,
CC thrombolytic, receptor/ligand, anti-inflammatory, cadherin/tumour
CC invasion suppressor, and tumour inhibition activities. The PNS are
CC also stated to be useful for gene therapy.

CC Sequence 481 BP; 185 A; 85 C; 90 G; 121 T; 0 other;

Query Match 100.0%; Score 481; DB 20; Length 481;
Best Local Similarity 100.0%; Pred. No. 2, 6e-98;
Matches 481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATACGTAAATAATAGAGACAGCTACAGTATCACTAATMAACCAAGGGATTTC 60
DB 1 ggaatcgtataataaagagagacagctacagtgatccaactaaacacagggatttc 60
QY 61 ATCAGACATTCCTCCGCTAATCATGAGTATTAATTAAGACGCAATTAATAA 120
DB 61 atcagacatctcccggtgtaatcagtgacagattataaagacgcaatgaattaa 120
QY 121 CATTTTACAGCTGCGGACACAACATCGAGTGGTCAATTTCTTCTGAAACGGTGTG 180
DB 121 cattttgacagctgcgcgacacaaactcgcagtggtcattttcttcgaaacggtgtg 180
QY 181 GTCCCTGAAAAGATGTTTCTCTGTTTTCATGAGCTGGCTGAAACCTGTGCATCAATAA 240
DB 181 gtccctgcaaaaagatgttctcgtttccatgagctgctgaaactgtcacatcaaa 240
QY 241 CAATACCCACATTTAGATGTTCAAGAAAGCCAGAGTAACCTATTCACGAATCA 300
DB 241 caataccacatltcagatgttcaagaagaagcagaaggtacatccatltccaagaatca 300
QY 301 AAAGAAATTAATTTGCTGTATAGAAAGTGATTCATGAGCAACCTGATTTTGTGTTTGTG 360
DB 301 aaagaataattgctgtatagaagtgtatcattcagacacactgattttgagtttgtg 360
QY 361 GAGCCGATGCTAAATAATTTGGAGCCAAAGACTCAAGATTAATGTAACCTGATCTCAAG 420
DB 361 gagccgatgctaaataatttggagccaaagactcaagaattaatgtaagctgatctcaag 420
QY 421 GCAAAATGCTATGTAATTTGAAAGCAAAAAAATAAATAAATAAATAAATAAATAA 480
DB 421 gcaaaatgctatgtaattttgaaagcaaaaaataaataaataaataaataaataa 480
QY 481 A 481

DB 481 a 481

RESULT 2

AAV88049
ID AAV88049 standard; cDNA; 318 BP.

AAV88049;

12-FEB-1999 (first entry)

EST clone FQ505.

Expressed sequence tag; secreted protein; haematopoiesis regulator;
tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolytic;
receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

Homo sapiens.

W09845437-A2.

15-OCT-1998.

10-APR-1998; 98WO-US06956.

10-APR-1997; 97US-0837312.

(GEMV) GENETICS INST INC.

Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
Racie LA, Spaulding V, Treacy M;

MP1: 1999-070078/06.

New polynucleotides encoding human secreted proteins - derived from
e.g. human blood, kidney, foetal lung, placenta, testes, brain,
ovary, pituitary, retina and colon cDNA libraries

Claim 1; Page 265; 641pp; English.

The present sequence represents an expressed sequence tag (EST), and is
a polynucleotide of the invention. The polynucleotides of the invention
are all secreted EST sequences isolated from a variety of human tissue
sources. The EST sequences and proteins encoded by them are predicted to
have useful biological activities which would make them suitable for
treating, preventing or ameliorating medical conditions in humans and
animals, although no supporting data is given. Suggested activities
include nutritional activity, immune stimulating or suppressing activity,
haematopoiesis regulating activity, tissue growth activity,
activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
and thrombolytic activity, receptor/ligand activity, anti-inflammatory
activity, cadherin/tumour invasion suppressor activity, tumour inhibition
activity. The EST sequences are also stated to be useful for gene
therapy.

Sequence 318 BP; 99 A; 64 C; 67 G; 88 T; 0 other;

Query Match 60.2%; Score 289, 6; DB 20; Length 318;
Best Local Similarity 98.6%; Pred. No. 1, 2e-55;
Matches 292; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 149 CGCAGGTGTTCAATTTCTTGTGAAACGGTGTGCTCCGTGCAAAAAGATGTTTCTGTGTTT 208
DB 22 cgcaagtggtcaattttcttcgaaaggtgtgtccctgcaaaagatgttctgtttt 81
QY 209 CCAAGAGCTGGCTGAAATTTCTCAGTCAATTAACCAATTTTCAAGTGTCAAGAA 268
DB 82 ccaagagctggtcgaatcttgcacatcaaaacatccacatltcagatgttcaagaa 141
QY 269 AAGCCAGAGTAACCTATTCACAGAAATCAAAAAGAAATTAATTTGCTTATAGAGTGG 328

Db	112	aagccagaagatgaacccattcttcgaataccaagaacaatttgcgtttatagaagtgg	201
Oy	329	ATTGATGAGCAACCGATTTTGTGAGTGCTATAAAAATTGGAGCCAA	368
Db	202	attcatgagcaaccgatatttgagittgtgagcgagatgctaanaaatltygaagc	261
Oy	369	GACTCAACAATTAAGTGAAGTCATCTCCAAGGCAAAATACACTTGTGCATTGGA	444
Db	262	gactcaagaatatgaatgaagctgaltctccaagcaaaatacaccttlygccgctcga	317
RESULT	3		
ID	AAA16169	standard; DNA; 631 BP.	
XX	AAA16169;		
DT	14-JUN-2000	(first entry)	
XX			
DE		Human colon cancer differentially expressed nucleotide sequence #174.	
KW		Colon cancer; detect; differential expression; human; treatment;	
KW		detect mutation; non-invasive diagnostic method; ds.	
OS		Homo sapiens.	
XX			
PN	WO200012702-A2.		
PD	09-MAR-2000.		
PF	30-AUG-1999;	99WO-US19424.	
PR	31-AUG-1998;	98US-0098639.	
PR	27-JAN-1999;	99US-0117393.	
PA	(FARB) BAYER CORP.		
P1	Endege WO,	Steinmann KE, Astle JH, Burgess CC, Carroll E;	
P1	Cacino TJ,	Dwivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;	
P1	Schlegel R;		
DR	WPI: 2000-256641/22.		
PT		Novel nucleic acids and proteins for identifying therapeutic agents	
PT		useful for treating and diagnosing cancer, especially colon cancer	-
PS			
XX	Claim 16; Page 193-194; 345pp; English.		
XX		This sequence represents a human nucleotide sequence which is	
CC		differentially expressed in colon cancer cells compared to the expression	
CC		levels in normal cells. The nucleotide sequence can be used as a source	
CC		of primers and probes. The nucleotide sequence is useful for determining	
CC		the phenotype of a cell by detecting the differential expression of the	
CC		sequence relative to a normal cell. The probes derived from the sequence	
CC		can also be used to determine the phenotype of cells in a sample. Probes	
CC		and antibodies which hybridise to the nucleotide sequence can also be	
CC		used to determine the phenotype of a cell. The primers are useful for	
CC		detecting a mutation in a test nucleotide sequence and also for detecting	
CC		cancer, preferably colon cancer. Antibodies against the protein encoded	
CC		by the nucleotide sequence can also be used in a method to detect colon	
CC		cancer. The diagnostic method is non-invasive and accurate for diagnosing	
CC		colon cancer at an early stage.	
XX			
SO	Sequence 631 BP; 190 A; 128 C; 136 G; 172 T; 5 other;		
Query Match	14.6%;	Score 70.4;	DB 21; Length 631;
Best Local Similarity	53.7%;	Pred. No. 9.-9e-07;	
Matches 191; Conservative	0;	Mismatches 161;	Indels 4; Gaps 2
Oy	62	TTCAGACTTCCTCCTGGTGTATCATCGTTACAGATTATTAAAGCACAGATGCAATTTAAC	121

Dd	116	tcttcaagcctccagcagccaagaatggtgaagaagtcgaaagacaagtgcctttccaaga	175
Oy	122	ATTTTTGAACAGCTGCCGGACACAACATCGCAGTGCTCAATTTCTTGGAAACGCTGCG	181
Dd	176	agccttgtagcgctgcagggtgatataacttgtaagtgtagctcttcacgccagctgggtcgg	235
Oy	182	TCCCTGCAAAAGGAGTGTTCCTCGTTTCCCATGTAGCTGGCTGTGAACACTTGTGCACATCAAAAC	241
Dd	236	gccttgcacaaaatgatcaaagccttcttcattccctctcctcttgaagaaglatcccaagctgat	295
Oy	242	AATACCACATTTTCAGANGTTCNAAGAAAAAGCCAGAAGTAACCTAATTTCTCAAGAATCAA	301
Dd	296	attccttgaagt--agatgltgatgacctgcagtcagatgltgtcctcaaaagtgtgaagttcaa	353
Oy	302	AAGAAATAATTTGCTGTATTAGAAAGTGATTCATGAGCAACCTGTAATTTTGTAGTTTTGGTG	361
Dd	354	atgcatt--gccacaccatccagtttctttaagaagaaggacaaaagtgtyggatattcttcgg	411
Oy	352	AGCCGATGCTAATAAATTTGGAAGCCCAAGCTCAAGAAATTAATGTAAGCTGATCCTCC	417
Dd	412	agccataaagaagaagcttgtaagcaccaccaattaatgaatcatcataatcatgttttc	467
RESULT 4			
AD23536			
ID	AD23536	standard; cDNA: 382 BP.	
XX	AD23536;		
DT	26-FEB-2002	(first entry)	
DE	Human lung tumour-specific 57212 CDNA.		
XX			
KW	Human; lung tumour protein; immunostimulant; cytostatic; gene therapy; antisense-therapy; vaccine; immune response; lung cancer; 57212; ss.		
OS	Homo sapiens.		
PN	WO200172295-A2.		
PD	04-OCT-2001.		
XX			
PF	28-MAR-2001; 2001WO-US09991.		
PR	29-MAR-2000; 2000US-0538037.		
PR	05-JUN-2000; 2000US-0588937.		
PR	18-AUG-2000; 2000US-0640878.		
PR	22-SEP-2000; 2000US-234517P.		
PR	01-NOV-2000; 2000US-070451Z.		
PR	14-DEC-2000; 2000US-0738973.		
PA	(CORI-) CORIXA CORP.		
P1	Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indrias CY;		
P1	Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;		
DR	WPI; 2001-639201/73.		
PT	New human lung-specific polynucleotides and polypeptides for the		
PT	diagnosis and treatment of disease e.g. lung cancer -		
XX			
PS	Claim 1; Page 358; 378pp; English.		
XX			
CC	The invention relates to isolated lung tumour-specific proteins and		
CC	their corresponding cDNA molecules. Lung tumour-specific proteins and		
CC	their antigen-presenting cells are useful for stimulating and/or		
CC	expanding T cells specific for a tumour protein, and for inhibiting		
CC	the development of cancer. The invention also relates to a composition		
CC	useful for stimulating an immune response, and for treating cancer. The		
CC	lung tumour specific oligonucleotide is useful in gene therapy and for		
CC	diagnosis, detection and treatment of lung cancer. The present sequence		
XX	is human lung tumour-specific cDNA.		

CC mainly to the 3' untranslated region (UTR) of the mRNA because they a

PD 18-JAN-1989.

Db 180 AGCCAAATGAAGAAAGCTTGACAGCCACCATTAATGATTTGCTAATCATGTTTTC 125

RESULT 9
AAA69777/c
ID AAA69777 standard; cDNA; 594 BP.

AC AAA69777;

DT 07-NOV-2000 (first entry)

DE Human ovarian carcinoma antigen polynucleotide SEQ ID NO:87.

XX Human: ovarian carcinoma; ovarian cancer; therapy; diagnosis;

XX tumour antigen; identification; cytostatic; gene therapy; vaccine; ss.

OS Homo sapiens.

PN WO200036107-A2.

PD 22-JUN-2000.

PF 17-DEC-1999; 99WO-US30270.

PR 17-DEC-1998; 98US-0215681.

PR 17-DEC-1998; 98US-0216003.

PR 23-JUN-1999; 99US-0338933.

PR 24-SEP-1999; 99US-0404879.

PA (CORI-) CORIXA CORP.

PI Mitcham JL, King GE, Algate PA, Frudakis TN;

DR WPI; 2000-431589/37.

PT Immunogenic portion of an ovarian carcinoma protein and the nucleic acid encoding it, useful for the diagnosis, prevention and treatment of cancer, preferably ovarian cancer -

PS Claim 18; Fig 15; 299pp; English.

XX The present invention describes an isolated polypeptide comprising an immunogenic portion of an ovarian carcinoma protein (or its variants).
XX Ovarian carcinoma proteins, and polynucleotides encoding them, have cytostatic activity and can be used in gene therapy and vaccines.
XX Ovarian carcinoma polypeptides, nucleic acids, antibodies and vaccines are useful for the prevention, diagnosis and treatment of cancer,
XX preferably ovarian cancer. AAA69691 to AAA70077 and AAB12552 to AAB12557 represent human ovarian carcinoma polynucleotides and proteins used in the exemplification of the present invention.

XX Sequence 594 BP; 169 A; 125 C; 114 G; 186 T; 0 other;

Query Match 12.5%; Score 60; DB 21; Length 594;

Best Local Similarity 52.0%; Pred. No. 0.00021; Matches 185; Conservative 0; Mismatches 165; Indels 6; Gaps 2;

QY 62 TCAGACCTCCCTGGTGAATCATGATGATTAAGACGAAATTTAAAC 121

Db 497 TCGTCAGACTCCAGACGCAAGATGTTGAGAGATGAGAGCAAGACTGTTTCAGGA 438

QY 122 ATTTTGGACGCGCCGGACAACTCGAGTGTTCATTTCTTCGAAACGGTGTGG 181

Db 437 AGCCTTGACGCTGACAGTGTATAACTTGTAGTACTTCCAGCCAGTGTGTGG 378

QY 182 TCCCTCAAAAGATGTTTCTGTTTTCATGAGCTGGTGAACCTGTGCATCAACAAAC 241

Db 377 GCCTTGCAAAATGATCAAGCTTCTTTCATTCCTCTCTGAAAGATTCACACGAT 318

QY 242 AATACCCACATTTTCAGATGTTTCAAGAAAGCCAGAAAGTAAACCTATTTCCAAAGATCAA 301

Db 317 ATTCCCTGAAGTAGATGTG---GATGACTGTGACAGATGTTGCTTCAGAGTGAAGTCAA 261

QY 302 AAGAAATATTTTGCCTGTTATAGAGTGCATTCATGACCACTGATTTTGTGTTGG 361

Db 260 ATGCATGCCAACAATCCAGTTTTTAAAG---AAGGACAAAAAGGTGGGTGAATTTCTGG 204

QY 362 AGCCGATGCTAAAAAATTTGGAAGCCACGACCAAGAAATTAATGTAAGCTGATCTCC 417

Db 203 AGCCAAATGAAGAAAGCTTGAGCCACCATTAATGATTAATGATGATGATGATTTTC 148

RESULT 10
AAA69823
ID AAA69823 standard; cDNA; 601 BP.

AC AAA69823;

DT 07-NOV-2000 (first entry)

DE Human ovarian carcinoma antigen polynucleotide SEQ ID NO:133.

XX Human: ovarian carcinoma; ovarian cancer; therapy; diagnosis;

XX tumour antigen; identification; cytostatic; gene therapy; vaccine; ss.

OS Homo sapiens.

PN WO200036107-A2.

PD 22-JUN-2000.

PF 17-DEC-1999; 99WO-US30270.

PR 17-DEC-1998; 98US-0215681.

PR 17-DEC-1998; 98US-0216003.

PR 23-JUN-1999; 99US-0338933.

PR 24-SEP-1999; 99US-0404879.

PA (CORI-) CORIXA CORP.

PI Mitcham JL, King GE, Algate PA, Frudakis TN;

DR WPI; 2000-431589/37.

PT Immunogenic portion of an ovarian carcinoma protein and the nucleic acid encoding it, useful for the diagnosis, prevention and treatment of cancer, preferably ovarian cancer -

PS Claim 18; Fig 15; 299pp; English.

XX The present invention describes an isolated polypeptide comprising an immunogenic portion of an ovarian carcinoma protein (or its variants).
XX Ovarian carcinoma proteins, and polynucleotides encoding them, have cytostatic activity and can be used in gene therapy and vaccines.
XX Ovarian carcinoma polypeptides, nucleic acids, antibodies and vaccines are useful for the prevention, diagnosis and treatment of cancer,
XX preferably ovarian cancer. AAA69691 to AAA70077 and AAB12552 to AAB12557 represent human ovarian carcinoma polynucleotides and proteins used in the exemplification of the present invention.

XX Sequence 601 BP; 172 A; 127 C; 138 G; 163 T; 1 other;

Query Match 12.5%; Score 60; DB 21; Length 601;

Best Local Similarity 52.0%; Pred. No. 0.00021; Matches 185; Conservative 0; Mismatches 165; Indels 6; Gaps 2;

QY 62 TCAGACCTCCCTGGTGAATCATGATGATTAAGACGAAATTTAAAC 121

Db 128 tcgtcagactccagcagcaagatggtgagcagatcgagcgcaagactgctttccagga 187

QY 122 ATTTTGGACGCGCCGGACAACTCGAGTGTTCATTTCTTCGAAACGGTGTGG 181

Db 188 agccttgacgctgacgagtgatgataactgtgtagtgcacttcagcagcagtggtg 247

CC Immunassay. The TC-CSF coding sequence, or portions of it, may be
 CC used in hybridisation assay or as primers.

XX Sequence 630 BP; 203 A; 102 C; 129 G; 196 T; 0 other;

Query Match 12.5%; Score 60; DB 14; Length 630;

Best Local Similarity 53.9%; Pred. No. 0.00021;
 Matches 193; Conservative 0; Mismatches 155; Indels 10; Gaps 3;

```

QY 62 TCAGCAGCTCCCTGGTGTATCATGATGATTTAAAGACAGCAATGATTTAAAC 121
    || || || || || || || || || || || || || || || || || || ||
Db 204 tcgcagactccagcaagatggtgagcagatcgagacagactgttttcagga 263
QY 122 ATTTTGACAGCTGCGGACAGCAACGCGTTCATTTCTTTCGAAACGCTGG 181
    || || || || || || || || || || || || || || || || || || ||
Db 264 agccttgagcgtcagtgataaactgtagtagtgaactccagccagctgtgtg 323
QY 182 TCCTGCAAAAGATGTTCTGTTTCATGAGCTGGTGAACCTGTCAATCAAAAC 241
    || || || || || || || || || || || || || || || || || || ||
Db 324 gacctgcaaaatgataagccttcttcattccctctcgaagaatctcaacgtgat 383
QY 242 AATACCCATTTCAGATGTTCAAGAAAGCCAGAGTAACCTTATTCAGAAATCAA 301
    || || || || || || || || || || || || || || || || || || ||
Db 384 atccctgaagtagatgtg--gatgactgtcagagatgtgtctcagagtgtagaagtc 440
QY 302 AAGAAATATTGCTGTATAGAGTGAATGATGAGCAACCTGATTTT--TGAGTTTGT 359
    || || || || || || || || || || || || || || || || || || ||
Db 441 atgca-----tgcacaactgtcagcttctttaaagaagagcaaaagtgtggaatttct 495
QY 360 GGAGCCGATGCTAAATAATTGAGCCAGACATCAAGATTAATTAATGATGATCTCC 417
    || || || || || || || || || || || || || || || || || || ||
Db 496 ggaagcaataagaaagcttgaagcaccattatgaattagctcaatcatgttttc 553

```

RESULT 13

AAD09895 AAD09895 standard; cDNA; 630 BP.

```

XX AC AAD09895;
XX DT 12-SEP-2001 (first entry)
XX DE Human T cell-derived colony stimulating factor (TC-CSF) cDNA.
XX KW T cell-derived colony stimulating factor; TC-CSF; immunostimulant;
KW acquired immune deficiency syndrome; AIDS; cancer; gene therapy;
KW haematopoietic growth factor; human; ss.
XX OS Homo sapiens.
XX FT Key
FH Location/Qualifiers
FT 1..630
FT CDS
FT /tag= a
FT /product= "Human TC-CSF protein"
FT /transl_except= (pos:541..543, aa:Xaa)
FT /transl_except= (pos:562..564, aa:Xaa)
FT /transl_except= (pos:628..630, aa:Xaa)
FT /note= "Xaa corresponds to in-frame stop codon;
FT CDS does not include start and stop codon"
FT /partial
FT sig_peptide
FT 55..84
FT /tag= b
FT mat_peptide
FT 85..627
FT /tag= c
FT /product= "Mature human TC-CSF protein"
XX US6254861-B1.
XX PN 03-JUL-2001.
XX PD 24-JAN-1994; 94US-0180371.
XX PF
XX OS

```

```

PR 23-MAY-1989; 89US-0356006.
PR 12-JUL-1991; 91US-0729135.
PR 19-AUG-1991; 91US-0747784.
PR 01-NOV-1991; 91US-0788115.
PR 01-JUL-1992; 92US-0906866.
XX
XX (CHOU/) CHOUDHURY C.
XX PA
XX PI
XX DR WPI; 2001-432041/46.
XX DR P-PSDB; AAE05136.
XX

```

Choudhury C;

WPI; 2001-432041/46.
 P-PSDB; AAE05136.

New T cell-derived colony stimulating factor for treating
 immune-compromised patients, including acquired immunodeficiency
 syndrome patients and certain types of cancer patients -
 Example 20; Column 47-50; 38pp; English.

The invention relates to T-cell derived colony stimulating factors
 (TC-CSF) and nucleic acid molecules encoding them. TC-CSF which
 is a novel haematopoietic growth factor is useful for treating
 immune-compromised patients, including acquired immune deficiency
 syndrome (AIDS) patients and certain types of cancer patients.
 TC-CSF DNA is used in gene therapy. The present sequence is human
 TC-CSF cDNA.

Sequence 630 BP; 202 A; 110 C; 123 G; 195 T; 0 other;

Query Match 12.5%; Score 60; DB 22; Length 630;

Best Local Similarity 52.0%; Pred. No. 0.00021;
 Matches 185; Conservative 0; Mismatches 165; Indels 6; Gaps 2;

```

QY 62 TCAGCAGCTCCCTGGTGTATCATGATGATTTAAAGACAGCAATGATTTAAAC 121
    || || || || || || || || || || || || || || || || || || ||
Db 204 tcgcagactccagcaagatggtgagcagatcgagacagactgttttcagga 263
QY 122 ATTTTGACAGCTGCGGACAGCAACGCGTTCATTTCTTTCGAAACGCTGG 181
    || || || || || || || || || || || || || || || || || || ||
Db 264 agccttgagcgtcagtgataaactgtagtagtgaactccagccagctgtgtg 323
QY 182 TCCTGCAAAAGATGTTCTGTTTCATGAGCTGGTGAACCTGTCAATCAAAAC 241
    || || || || || || || || || || || || || || || || || || ||
Db 324 gacctgcaaaatgataagccttcttcattccctctcgaagaatctcaacgtgat 383
QY 242 AATACCCATTTCAGATGTTCAAGAAAGCCAGAGTAACCTTATTCAGAAATCAA 301
    || || || || || || || || || || || || || || || || || || ||
Db 384 atccctgaagtagatgtg--gatgactgtcagagatgtgtctcagagtgtagaagtc 440
QY 302 AAGAAATATTGCTGTATAGAGTGAATGATGAGCAACCTGATTTT--TGAGTTTGT 361
    || || || || || || || || || || || || || || || || || || ||
Db 441 atgcatgccaacatccagcttctttaaag--aagggacaaaagtgtggaattctcgg 497
QY 362 AGCGATGCTTAAATAATTGAGCCAGACATCAAGATTAATTAATGATGATCTCC 417
    || || || || || || || || || || || || || || || || || || ||
Db 498 agccaataagaaagcttgaagcaccattatgaattagctcaatcatgttttc 553

```

RESULT 14

AAQ49402 AAQ49402 standard; DNA; 456 BP.

AAQ49402;

04-MAY-1994 (first entry)

14kDa ECEF/thioredoxin/ADF coding sequence.

Eosinophil cytotoxicity enhancing factor; ECEF; stimulation;
 diagnosis; therapy; thioredoxin; ss.
 Homo sapiens.

Sat Jun 8 10:14:10 2002

us-09-746-783-18.rng

Page 11

Search completed: June 8, 2002, 03:25:51
Job time: 4131 sec

PD 6/5/98

p 122

Sat Jun 8 10:14:12 2002

us-09-746-783-19.rag

SEARCH ID NO: 19
Database: A-Geneseg-032802
AC NO: AAW95351

File Copy
Page 11

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2002, 03:20:36 ; Search time 61.07 Seconds
(without alignments)
194.611 Million cell updates/sec

Title: US-09-746-783-19

Perfect score: 562
Sequence: 1 MWOLIKDNEFKETLTAAGH.....IFFECGADAKKLEANTQELM 107

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

A-Geneseg-032802: *
1: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1980.DAT: *
2: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1981.DAT: *
3: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1982.DAT: *
4: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1983.DAT: *
5: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1984.DAT: *
6: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1985.DAT: *
7: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1986.DAT: *
8: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1987.DAT: *
9: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1988.DAT: *
10: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1989.DAT: *
11: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1990.DAT: *
12: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1991.DAT: *
13: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1992.DAT: *
14: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1993.DAT: *
15: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1994.DAT: *
16: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1995.DAT: *
17: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1996.DAT: *
18: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1997.DAT: *
19: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1998.DAT: *
20: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA1999.DAT: *
21: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA2000.DAT: *
22: /SIDSI/gcgdata/geneseg/genesegp-emb1/AA2001.DAT: *

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	562	100.0	107	20	AAW95351
2	190	33.8	105	10	AAW92141
3	190	33.8	105	14	AAW92141
4	190	33.8	105	16	AAW92141
5	190	33.8	105	20	AAW92141
6	190	33.8	105	21	AAW92141
7	190	33.8	105	22	AAW92141
8	189	33.6	105	15	AAW92141
9	187	33.3	105	14	AAW92141
10	185	32.9	104	15	AAW92141
11	185	32.9	104	16	AAW92141

12	184	32.7	172	10	AAW94809	Human Interleukin
13	184	32.7	209	14	AAW94809	Human TC-CSF, Hom
14	183	32.6	104	14	AAW94809	Human TC-CSF, Hom
15	180	32.0	105	19	AAW94809	Human thioresoxin
16	179	31.9	105	19	AAW94809	Human thioresoxin
17	168	29.9	105	19	AAW94809	Human thioresoxin
18	155	27.6	84	14	AAW94809	Human thioresoxin
19	140	24.9	80	14	AAW94809	Human thioresoxin
20	134	23.8	287	22	AAW94809	Human thioresoxin
21	118	21.0	130	19	AAW94809	Human thioresoxin
22	117.5	20.9	106	22	AAW94809	Human thioresoxin
23	111	19.8	102	20	AAW94809	Human thioresoxin
24	101.5	18.1	107	22	AAW94809	Human thioresoxin
25	100.5	17.9	133	21	AAW94809	Human thioresoxin
26	100.5	17.9	134	21	AAW94809	Human thioresoxin
27	98.5	17.5	178	21	AAW94809	Human thioresoxin
28	95	16.9	144	21	AAW94809	Human thioresoxin
29	94	16.7	122	21	AAW94809	Human thioresoxin
30	93.5	16.6	177	21	AAW94809	Human thioresoxin
31	93.5	16.6	177	21	AAW94809	Human thioresoxin
32	93.5	16.6	275	21	AAW94809	Human thioresoxin
33	93.5	16.6	275	21	AAW94809	Human thioresoxin
34	92.5	16.5	123	22	AAW94809	Human thioresoxin
35	91	16.2	126	21	AAW94809	Human thioresoxin
36	91	16.2	127	21	AAW94809	Human thioresoxin
37	91	16.2	139	21	AAW94809	Human thioresoxin
38	91	16.2	139	21	AAW94809	Human thioresoxin
39	91	16.2	221	21	AAW94809	Human thioresoxin
40	91	16.2	228	21	AAW94809	Human thioresoxin
41	90.5	16.1	115	21	AAW94809	Human thioresoxin
42	90.5	16.1	122	15	AAW94809	Human thioresoxin
43	90.5	16.1	122	15	AAW94809	Human thioresoxin
44	89.5	15.9	127	21	AAW94809	Human thioresoxin
45	89	15.8	109	17	AAW93015	Thioresoxin h cons

ALIGNMENTS

RESULT 1	
AAW95351	standard; Protein: 107 AA.
ID	AAW95351
AC	AAW95351
DT	26-APR-1999 (first entry)
DE	Human adult testis secreted protein fg505_4.
XX	Secreted protein; human; testis; fg505_4.
OS	Homo sapiens.
XX	
PN	W09856909-A2.
XX	
PD	17-DEC-1998.
XX	
PF	08-JUN-1998; 98WO-US11822.
XX	
PR	05-JUN-1998; 98US-0092722.
PR	11-JUN-1997; 97US-0873218.
XX	
PA	(GENY) GENETICS INST INC.
XX	
PI	Agostino MJ, Recheil K, Howes SH, Jacobs K, Lavallie ER.
PI	McCoy JM, Racie LA, Spaulding V, Treacy M;
XX	
DR	WPI: 1999-080899/07.
XX	
DR	N-PSDB: AAW99731.
XX	
PT	New polynucleotides encoding secreted human proteins - derived from
PT	human foetal brain, adult testes, foetal kidney, adult thyroid or
PT	adult retina cDNA libraries

Query Match	100.0%;	Score 562;	DB 20;	Length 107;
Best Local Similarity	100.0%;	Pred. No. 6.2e-64;		
Matches 107; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

RESULT	2
.AAP92141	
ID AAP92141 standard; protein; 105 AA	

XX	AA92141:	
AC	25-JUN-1990	(first entry)
DT		
XX		
DE	Recombinant human adult T cell leukemia derived factor.	
XX		
XX	Human adult T cell leukemia derived factor; HADP; cancer;	
KW	immunodeficiency disease.	
XX		
OS	Homo sapiens.	
XX		
XX	Key	Location/Qualifiers
FT	Region	2..105
FT		/note="Claim 1"
XX		
PN	EP299206-A.	
XX		
PD	18-JAN-1989.	
XX		
PF	10-JUN-1988;	88EP-0109311.
XX		
PR	12-JUN-1987;	87JP-0146348.
XX		
PA	(AJIN) AJINOMOTO KK.	
XX		
PI	Yodol J, Tagaya Y, Maeda M, Matsui H, Kondo N, Hamuro J;	
XX		
WI	WI; 1989-016762/03.	
DR	N-PSDB; AAN93083.	
XX		
XX	Recombinant human adult T cell leukemia derived factor polypeptide -	
XX	used for treating cancer, immunodeficiency disease etc.	

Query Match	33.8%	Score 190;	DB 10;	Length 105;
Best Local Similarity	36.2%	Pred. No. 2.1e-16;		
Matches	46;	Conservative 11;	Mismatches 28;	Indels 42; Gaps 2.

RESULT 3
AAR37700
ID AAR37700 standard; Protein; 105 AA

DT 09-NOV-1993 (first entry)
 XX
 DE ADF.
 XX
 XX Human; ADF; transgenic; mouse; beta-actin; promoter; terminator;
 KM resistant; stress; anti-inflammatory drugs.
 XX
 OS Homo sapiens.
 XX
 PN JP05130819-A.
 XX
 PD 28-MAY-1993.
 XX
 PF 12-NOV-1991; 91JP-0295618.
 XX
 PR 12-NOV-1991; 91JP-0295618.
 XX
 PA (AJIN) AJINOMOTO KK.
 XX
 DR WPI; 1993-208254/26.
 DR N-PSDB; AAQ43433.
 XX
 PT New human ADF transgenic mouse - is resistant to stress and is
 PT useful for prepn. of antiinflammatory drugs
 XX
 PS Claim 3; Page 5; 10pp; Japanese.
 PS
 CC This sequence represents human ADF. The DNA encoding this sequence
 CC may be used in the production of a transgenic mouse. The transgenic
 CC mouse contains, in the 5' to 3' direction, a human beta-actin promoter,
 CC the human ADF coding gene, a termination codon, a human beta-actin
 CC poly(A) signal and a human beta-actin terminator. The transgenic mouse
 CC is resistant to stress. It is useful in the research of the mechanism
 CC of stress and for the development of anti-inflammatory drugs.
 XX

Sat Jun 8 10:14:10 2002

us-09-746-783-18.rmg

SEA ID NO: 18
Database: N-Geneseg-032802
AC, NO: AAV99731

File copy
Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2002, 02:17:00 ; Search time 272.98 Seconds
(without alignments)
3025.261 Million cell updates/sec

Title: US-09-746-783-18

Perfect score: 481
Sequence: 1 GGATACCTGTAATAATAGCA.....AAAAAAAAAAAAAAAAAAAA 481

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

N.Geneseg_032802: *
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: *
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: *
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: *
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: *
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: *
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	481	100.0	481	20	AAV99731
2	289.6	60.2	318	20	AAV88049
3	70.4	14.6	631	21	AAV16169
4	61.6	12.8	382	23	AAV23536
5	61.6	12.8	516	21	AAV3962
6	60.6	12.6	581	10	AAV3083
7	60.6	12.6	581	20	AAV20054
8	60	12.5	557	21	AAV69778
9	60	12.5	594	21	AAV69777

ALIGNMENTS

RESULT	ID	AAV99731 standard; cDNA; 481 BP.	ALIGNMENTS	Human ovarian carc.
1	AAV99731			
AC	AAV99731			Human colon cancer
AC	AAV99731			Human TC-CSF, Hom
AC	AAV99731			Human T cell-deriv
AC	AAV99731			14kDa ECFR/thored
AC	AAV99731			ADP gene, Homo sa
AC	AAV99731			Sequence encoding
AC	AAV99731			Human ADF-polypept
AC	AAV99731			Human colon cancer
AC	AAV99731			RECF 104 coding s
AC	AAV99731			Human ADF-polypept
AC	AAV99731			Sequence encoding
AC	AAV99731			RECF 84 coding se
AC	AAV99731			Seq ID 29 from del
AC	AAV99731			cDNA sequence #26
AC	AAV99731			Human immune syste
AC	AAV99731			Tumour suppressor
AC	AAV99731			T cell replacing f
AC	AAV99731			Plasmod psp6K-mTRF
AC	AAV99731			Tumour suppressor
AC	AAV99731			Human immune syste
AC	AAV99731			Human cervical can
AC	AAV99731			Human secreted pro
AC	AAV99731			Human bladder tumo
AC	AAV99731			Human secreted pro
AC	AAV99731			Human breast cance
AC	AAV99731			Human cervical can
AC	AAV99731			Human breast cance
AC	AAV99731			Human breast cance
AC	AAV99731			Human immunodefici
AC	AAV99731			cDNA encoding nove
AC	AAV99731			Human secreted pro
AC	AAV99731			Human secreted pro
AC	AAV99731			Human immune syste
AC	AAV99731			Human polynucleoti

DR WPI: 1999-080899/07.
 DR P-PSDB: AAW95351.
 XX New polynucleotides encoding secreted human proteins - derived from
 PT human foetal brain, adult testes, foetal kidney, adult thyroid or
 PT adult retina cDNA libraries
 PS
 XX Claim 30; Page 85; 113pp; English.

CC This is the nucleotide sequence of cDNA clone fg505_4, which
 CC includes an open reading frame for a 107-amino acid polypeptide
 CC (see AAW95351). The clone was isolated from a human adult testis
 CC cDNA library using methods which are selective for cDNAs encoding
 CC secreted proteins, or was identified as encoding a secreted or
 CC transmembrane protein on the basis of computer analysis of the
 CC amino acid sequence of the encoding protein. Database searches
 CC indicate some sequence similarity to known sequences. The
 CC invention provides cDNA clones (see AAW9721-33) from human adult
 CC thyroid, adult retina, adult testis, foetal kidney and foetal brain
 CC that encode novel secreted proteins (see AAW95344-53). Each clone is
 CC individually available from deposit clone ATCC 98451 (see also
 CC AAW9734-43). The isolated polynucleotides (Pns) and proteins are
 CC predicted to have activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans
 CC and animals, although no supporting data is given. Suggested
 CC activities include nutritional, cytokine, cell proliferation or
 CC differentiation, immune stimulating (e.g. as vaccines) or immune
 CC suppressing, haematopoiesis regulating, tissue growth,
 CC activin/inhibin, chemotactic/chemokinetic, haemostatic,
 CC thrombolytic, receptor/ligand, anti-inflammatory, cadherin/tumour
 CC invasion suppressor, and tumour inhibition activities. The Pns are
 CC also stated to be useful for gene therapy.

XX Sequence 481 BP; 185 A; 85 C; 90 G; 121 T; 0 other;

Query Match 100.0%; Score 481; DB 20; Length 481;
 Best Local Similarity 100.0%; Pred. No. 2.6e-98;
 Matches 481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 1 GGATCTGATTAATATAGAGACAGCTACAGTATCCAACTAAGCAGGAGATTTC 60
 DB 1 ggatactgtaataataatagagacagctacagatccaactaacacagggatcttc 60
 0Y ATCAGACACTTCCCTGGTGTATCATGATGATCAATTTAAAGACAGCAATGAATTTAAA 120
 DB 61 atcagacacttccctgggtgatcatgatactgatacttaagacagcaatgaatttaaa 120
 0Y 121 CATTTTGACAGTGGCGGACAACTGCGAGTGTCAATTTCTTGAACGGTGTG 180
 DB 121 catlttgacagtcgcccgcagacaaactcgagtggtcaatttctcgaaacggtgtg 180
 0Y 181 GTCCCTGCAAAAGATGTTTCTGTTTCCATGAGCTGGCTGAACTTGTCACATCAAAA 240
 DB 181 gtccctgcaaaagatggttctgtttccatgagctggtcgaactgttcacataaaa 240
 0Y 241 CAATACCCCATTTTCAGATGTTTCAGAAAAGCCAGAGTAACCTTAATTTCTCAAGATCA 300
 DB 241 caatacccatcttcagatgttcaagaagaagcagaagtaaccctattctcaagatca 300
 0Y 301 AAGAGTAATTTGCTGTTTATAGAGTATCATGAGCAACGATTTTGATTTGTG 360
 DB 301 aagagtaatttgcgtgttataagaagtgatcagcaaccgattttgagttgtg 360
 0Y 361 GAGCGAGTCTAAAAAATTTGGAAGCCAGACTCAAGATTAATTAAGTGTATCTCAAG 420
 DB 361 gagcgagtgctaaaaaatttggaagccaagctcaagatataatgtaagctgtctcaag 420
 0Y 421 GCAAAATFACCTTTGTGACATTTGAAAAGCAAAAAAAGAAAAAAGAAAAAAGAAAA 480
 DB 421 gcaaaatfacctttgtgacatTTTgaaaagcAAAAAAGAAAAAAGAAAAAAGAAAA 480
 0Y 481 A 481

DB 481 A 481

RESULT 2
 ID AAW88049 standard; cDNA; 318 BP.
 AC AAW88049;
 XX
 AC AAW88049;
 XX
 DT 18-FEB-1999 (first entry)
 XX
 DE EST clone FQ505.
 DE
 XX
 XX Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 XX
 OS Homo sapiens.
 XX
 PN MO9845437-42.
 XX
 PD 15-OCT-1998.
 XX
 PF 10-APR-1998; 98WO-US06956.
 XX
 PR 10-APR-1997; 97US-0837312.
 XX
 PA (GENET) GENETICS INST INC.
 XX
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racie LA, Spaulding V, Treacy M;
 XX
 DR WPI: 1999-070078/06.

PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries
 XX
 XX Claim 1; Page 265; 641pp; English.

CC The present sequence represents an expressed sequence tag (EST), and is
 CC a polynucleotide of the invention. The polynucleotides of the invention
 CC are all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The EST sequences are also stated to be useful for gene
 CC therapy.

SO Sequence 318 BP; 99 A; 64 C; 67 G; 88 T; 0 other;

Query Match 60.2%; Score 289.6; DB 20; Length 318;
 Best Local Similarity 98.6%; Pred. No. 1.2e-55;
 Matches 292; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

0Y 149 GCGAGTGTTCAATTTTCTTCAAGAGGTGTGCTCCCTGCAAGAGATGTTCTTCTTT 208
 DB 22 gcgagtggtcaatttcttcgaagcgtgtgtctccctgcaagagatgttctctgttt 81
 0Y 209 CCATGAGCTGGCTGAAATTTGTCACATCAAAACATTAACCATTTTCAGATGTTCAAGAA 268
 DB 82 ccatagctgctgtaactgttcacatcaacaataccacatcttcagatgttcaagaa 141
 0Y 269 AAGCAGAGAGGTAAACCTTAATTTCTCAAGATCAAAAGAAATTAATTTGCTGTATGAAGTGC 328


```

CGAGCTGGGCTTGGGTTACAGAGCCGAGACCCCTGAGACTCCGCCACGCTTGCCAGACCCCGCCGA
470      480      490      500      510      520      530
      30      40      50      60      70      80      90
TTTTTTTTGCTTTTCAANGTCACAGTGTATTTTGCCTTGAGATCAGCTTACATTAAATCTTGAGCTT
      100      110      120      130      140      150      160
TCCGGGCTCGAGGCTCCAGGGGGGCGGTGCGCGGGGCGCAGCGAGCGGGCGAGTCCGGCGGGCGC
540      550      560      570      580      590      600
TGCTTCCAAATTTTTAGCATCGGCTCCACAAAACCTCAAAAATCAGGTTGCTCATGAATCCACTTAAC
      170      180      190      200      210      220      230
GCGGCGCCCGCCACAGCGCGCGCGGCGGCTCCAGTTTATTAAGGGAGAGAGCAGCAGCGAGCTTGAAGC
610      620      630      640      650      660      670
AGCAAAATTTCTTTTGTGATCTTGAGAAATAGG-GTTACCTTCTGCTTTTCTTGAACATCTGAAATGTGGGT
      170      180      190      200      210      220      230
TCTGTTGTGTCCTTGGATCCATTTCCATTCGCTCCTTACAGCCGCTCTCAGACTCCAGACCAAGATGTT
680      690      700      710      720      730      740      750
      240      250      260      270      280      290      300
AATTGTTGATGTGACAAAGTTTCAGCCAGCTTCATGAAAAACAGAAACATCCCTTTTGCAAGGACACACCGT
      310      320      330
GAAGCAGATCGAGCAGGTACGCGCTACCGGGGAGAGCGCAGGCGTCCGC
760      770      780      790      800
TTGGAAGAAAAATGAACCATGCGAGTTT
310      320      330

```


GeneCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2002, 03:20:36 ; Search time 61.07 Seconds

(without alignments)
194.611 Million cell updates/sec

Title: US-09-746-783-19

Perfect score: 562
Sequence: 1 MVQIKDTNEFKFTLTAAGH.....IFEFCGADAKKLEAKTQELM 107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A-GeneSeq_032802:*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	562	100.0	107	AAW95351
2	190	33.8	105	AAW92141
3	190	33.8	105	AAW97700
4	190	33.8	105	AAW92390
5	190	33.8	105	AAW95908
6	190	33.8	105	AAW93956
7	190	33.8	210	AAW95136
8	189	33.6	105	AAW97856
9	187	33.3	105	AAW92819
10	185	32.9	104	AAW97853
11	185	32.9	104	AAW92389

12	184	32.7	172	10	AAW94809
13	184	32.7	209	14	AAW93191
14	183	32.6	104	14	AAW942820
15	180	32.0	105	19	AAW959567
16	179	31.9	105	19	AAW959566
17	168	29.9	105	19	AAW959565
18	155	27.6	84	14	AAW942821
19	140	24.9	80	14	AAW942822
20	134	23.8	287	22	AAW961581
21	118	21.0	130	19	AAW970535
22	117.5	20.9	106	22	AAW960229
23	111	19.8	102	20	AAW935279
24	101.5	18.1	107	22	AAW960478
25	100.5	17.9	133	21	AAW910853
26	100.5	17.9	134	21	AAW904653
27	98.5	17.5	178	21	AAW919366
28	95	16.9	144	21	AAW934790
29	94	16.7	122	21	AAW970479
30	93.5	16.6	177	21	AAW916557
31	93.5	16.6	177	21	AAW942782
32	93.5	16.6	275	21	AAW916556
33	93.5	16.6	275	21	AAW942781
34	92.5	16.5	123	22	AAW937102
35	91	16.2	126	21	AAW920450
36	91	16.2	127	21	AAW924028
37	91	16.2	127	21	AAW940198
38	91	16.2	139	21	AAW920449
39	91	16.2	221	21	AAW940197
40	91	16.2	228	21	AAW940196
41	90.5	16.1	115	21	AAW912953
42	90.5	16.1	122	15	AAW95908
43	90.5	16.1	122	21	AAW935810
44	89.5	15.9	127	21	AAW926604
45	89	15.8	109	17	AAW93015

ALIGNMENTS

RESULT 1	AAW95351	standard; Protein; 107 AA.
XX	AAW95351;	
AC	26-APR-1999	(first entry)
DT	Human adult testis secreted protein fg505_4.	
DE	Secreted protein; human; testis; fg505_4.	
XX	Human sapiens.	
OS	Homo sapiens.	
XX	WO9856909-A2.	
PN	17-DEC-1998.	
XX	08-JUN-1998;	98WO-US11822.
PF	05-JUN-1998;	98US-0092722.
XX	PR 11-JUN-1997;	97US-0873218.
XX	(GENY) GENETICS INST INC.	
PA	Agostino MJ, Fechtel K, Howes SH, Jacobs K, LaVallie ER;	
XX	McCoy JM, Racie LA, Spaulding V, Treacy M;	
PI	WPI; 1999-080899/07.	
XX	N-PSDB; AAV99731.	
DR	New polynucleotides encoding secreted human proteins - derived from	
XX	human foetal brain, adult testes, foetal kidney, adult thyroid or	
PT	adult retina cDNA libraries	

XX Claim 31; Page 86; 113pp; English.
 PS
 XX
 CC This is the amino acid sequence of fg505.4, a novel human
 CC secreted protein predicted from the nucleotide sequence of a
 CC human adult testis cDNA clone (see AAV99731). Database searches
 CC indicate some sequence similarity to known sequences. The
 CC invention provides cDNA clones (see AAV99721-33) from human adult
 CC thyroid, adult retina, adult testis, foetal kidney and foetal
 CC brain that encode novel secreted proteins (see AAV95344-53). The
 CC polynucleotides and proteins are predicted to have activities
 CC which would make them suitable for treating, preventing or
 CC ameliorating medical conditions in humans and animals, although no
 CC supporting data is given. Suggested activities include nutritional,
 CC cytokine, cell proliferation/differentiation, immune stimulating
 CC (e.g. as vaccines) or immune suppressing, haematopoiesis regulating,
 CC tissue growth, activin/inhibin, chemotactic/chemokinetic,
 CC haemostatic, thrombolytic, receptor/ligand, antiinflammatory,
 CC cadherin/tumour invasion suppressor, and tumour inhibition
 CC activities.
 CC
 CC Sequence 107 AA;
 XX
 SQ

Query Match 100.0%; Score 562; DB 20; Length 107;
 Best Local Similarity 100.0%; Pred. No. 6.2e-64;
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MWQIKDNEFTFTLAGHKLAVVQSSKRCGPCRMFPVPHLAECHIKITPFQWF 60
 |||
 DB 1 mwqikdneftftlaaghklaavvqsskrcgpcrmpvphlaechikitpfqwf 60
 OY 61 KKSQVTLFSRIKRIICCYRSGFMSNLIFFPCGADAKKLEAKTOELM 107
 |||
 DB 61 kksqvtlfsrikriiccyrgsfmsnliffpcgadakkleaktelgm 107

RESULT 2

AAP92141
 ID AAP92141 standard; protein; 105 AA.

AC AAP92141;

DT 25-JUN-1990 (first entry)

DE Recombinant human adult T cell leukaemia derived factor.

KW Human adult T cell leukaemia derived factor; hADF; cancer;

KW Immunodeficiency disease.

OS Homo sapiens.

FX Key Location/Qualifiers

FT Region 2..105

FT /note= "Claim 1"

XX EP299206-A.

XX 18-JAN-1989.

XX 10-JUN-1988; 88EP-0109311.

XX 12-JUN-1987; 87JP-0146348.

XX (AJIN) AJINOMOTO KK.

XX Yodoi J, Tagaya Y, Maeda M, Matsui H, Kondo N, Hamuro J;

XX WPI; 1989-016762/03.

XX DR N-PSDB; AAN93083.

XX Recombinant human adult T cell leukaemia derived factor polypeptide -

PT used for treating cancer, immunodeficiency disease etc.

XX Disclosure; 24pp; English.
 PS
 XX
 CC DNA encoding the polypeptide was sequenced from a gene bank prep. from
 CC mRNA isolated from ATL-2 cells from patient with adult T leukaemia
 CC virus. Vectors contg. the DNA can be used to transform host cells for
 CC prodn. of hADF polypeptide. The polypeptide causes differentiation and
 CC induces growth of lymphocytes and fibroblasts. The N-terminal Met is
 CC optional. Similar peptides with deletions, sustns. and allelic derivs.
 CC may also be used.
 CC See also AAP94809.
 CC
 CC Sequence 105 AA;
 XX
 SQ

Query Match 33.8%; Score 190; DB 10; Length 105;
 Best Local Similarity 36.2%; Pred. No. 2.1e-16;
 Matches 46; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

OY 1 MWQIKDNEFTFTLAGHKLAVVQSSKRCGPCRMFPVPHLAE----- 47
 |||
 DB 1 mwqikdneftftlaaghklaavvqsskrcgpcrmpvphlae----- 47
 OY 48 -----TCHIKITPFQWFKSKOKVTLFSRIKRIICCYRSGFMSNLIFFPCGADAKKLE 100
 |||
 DB 61 dcqdvasecevcmpbfqfkkqkvq-----eisganxkle 98
 OY 101 AKTOELM 107
 |||
 DB 99 atlneliv 105

RESULT 3

AAR37700
 ID AAR37700 standard; protein; 105 AA.

AC AAR37700;

DT 09-NOV-1993 (first entry)

DE ADF.

KW Human; ADF; transgenic; mouse; beta-actin; promoter; terminator;

KW resistant; stress; anti-inflammatory drugs.

OS Homo sapiens.

PN JP05130819-A.

PD 28-MAY-1993.

PD 12-NOV-1991; 91JP-0295618.

PR 12-NOV-1991; 91JP-0295618.

XX (AJIN) AJINOMOTO KK.

XX WPI; 1993-208254/26.

XX DR N-PSDB; AAO43433.

XX New human ADF transgenic mouse - is resistant to stress and is

XX useful for prepn. of antiinflammatory drugs

XX Claim 3; Page 5; 10pp; Japanese.

CC This sequence represents human ADF. The DNA encoding this sequence
 CC may be used in the production of a transgenic mouse. The transgenic
 CC mouse contains, in the 5' to 3' direction, a human beta-actin promoter,
 CC the human ADF coding gene, a termination codon, a human beta-actin
 CC poly(A) signal and a human beta-actin terminator. The transgenic mouse
 CC is resistant to stress. It is useful in the research of the mechanism
 CC of stress and for the development of anti-inflammatory drugs.

SQ Sequence 105 AA;

Query Match 33.8%; Score 190; DB 14; Length 105;
 Best Local Similarity 36.2%; Pred. No. 2.1e-16;
 Matches 46; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

QY 1 MVOIIKDTEFFKTELTAAAGHKLAVVOPSSKRCGCPCKRMFPVHELAE-----47

DB 1 mvqiesktafgealdaagdklvvdtsatwcpckmkpffhslseksynviflevdvd 60

QY 48 -----TCHIKTIPTEFQMEKRSOKVTLFSRIKRIICCYNSGFSNLIFFECGADAKKLE 100

DB 61 dcqdvasecevkcmptfqqfkkqgkv-----efsgankekile 98

QY 101 AKTOELM 107

DB 99 atlnelv 105

RESULT 4

ID AAR72390 standard; Protein; 105 AA.

AC AAR72390;

DT 10-NOV-1995 (first entry)

DE Recombinant human ADF.

KW ADF; Inflammation; radiation sickness; fermentation; recombinant.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Active-site 32..35

PN JP07079780-A.

PD 28-MAR-1995.

PF 20-SEP-1993; 93JP-0233361.

PR 20-SEP-1993; 93JP-0233361.

PA (AJIN) AJINOMOTO KK.

DR WPI; 1995-157850/21.

DR N-PSDB; AAQ87818.

XX Prepn. of recombinant human ADF by direct expression in E. coli -
 PT for treating inflammation and radiation sickness caused by
 PT commercially produced free radical(s)

PS Claim 3; Page 8; 11pp; Japanese.

CC The DNA sequence encoding human ADF was inserted into a plasmid which
 CC was then used to transform E.coli. The E.coli was then cultured and
 CC the recombinantly produced ADF harvested. Human ADF can be used as a
 CC treating agent for inflammation and radiation sickness.

SQ Sequence 105 AA;

Query Match 33.8%; Score 190; DB 16; Length 105;
 Best Local Similarity 36.2%; Pred. No. 2.1e-16;
 Matches 46; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

QY 1 MVOIIKDTEFFKTELTAAAGHKLAVVOPSSKRCGCPCKRMFPVHELAE-----47

DB 1 mvqiesktafgealdaagdklvvdtsatwcpckmkpffhslseksynviflevdvd 60

QY 48 -----TCHIKTIPTEFQMEKRSOKVTLFSRIKRIICCYNSGFSNLIFFECGADAKKLE 100

DB 61 dcqdvasecevkcmptfqqfkkqgkv-----efsgankekile 98
 QY 101 AKTOELM 107
 DB 99 atlnelv 105

RESULT 5

ID AAY25908 standard; Protein; 105 AA.

AC AAY25908;

DT 06-OCT-1999 (first entry)

DE Human thioredoxin protein.

XX Thioredoxin; thioredoxin reductase; human; antisense; primer; metastasis;
 KW cytosolic; tumour growth inhibitor; detection; nuclease resistant;
 KW phosphorothioate linkage.

OS Homo sapiens.

PN WO9338963-A1.

PD 05-AUG-1999.

PF 29-JAN-1999; 99WO-CA00077.

PR 30-JAN-1998; 98US-0073196.

PA (GENE-) GENESENSE TECHNOLOGIES INC.

PI Lee YS, Wright JA, Young AH;

DR WPI; 1999-469328/39.

DR N-PSDB; AAZ00544.

PT Antisense oligonucleotides against thioredoxin and thioredoxin

PT reductase genes, useful for inhibiting tumor growth and metastasis
 XX Disclosure; Fig 2; 88pp; English.

PS This invention describes novel antisense oligonucleotides against

CC thioredoxin and thioredoxin reductase gene which have cytostatic
 CC activity and are useful for inhibiting tumor growth and metastasis
 CC in mammals. They may also be used as hybridization probes to detect
 CC the presence of the thioredoxin and thioredoxin reductase mRNAs in
 CC mammalian cells. They may also be used as molecular weight markers.

CC The antisense oligonucleotides are nuclease resistant due to the
 CC presence of phosphorothioate internucleotide linkages. This sequence
 CC represents the human thioredoxin protein.

XX Sequence 105 AA;

Query Match 33.8%; Score 190; DB 20; Length 105;
 Best Local Similarity 36.2%; Pred. No. 2.1e-16;
 Matches 46; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

QY 1 MVOIIKDTEFFKTELTAAAGHKLAVVOPSSKRCGCPCKRMFPVHELAE-----47

DB 1 mvqiesktafgealdaagdklvvdtsatwcpckmkpffhslseksynviflevdvd 60

QY 48 -----TCHIKTIPTEFQMEKRSOKVTLFSRIKRIICCYNSGFSNLIFFECGADAKKLE 100

DB 61 dcqdvasecevkcmptfqqfkkqgkv-----efsgankekile 98

QY 101 AKTOELM 107

DB 99 atlnelv 105

```

RESULT 6
AAG03956
ID AAG03956 standard; Protein: 105 AA.
XX
AC AAG03956;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 8037.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
OS Homo sapiens.
XX
PN EPI033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GIST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI: 2000-500381/45.
DR N-PSDB: AAC03962.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 13; SEQ ID 8037; 71pp + CD-ROM; English.
XX
XX
XX The present sequence is a polypeptide encoded by one of a large number
XX of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
XX were prepared from total human RNAs or polyA+ RNAs derived from 30
XX different tissues. EST sequences usually correspond mainly to the 3'
XX untranslated region (UTR) of the mRNA because they are often obtained
XX from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
XX isolating cDNA sequences derived from the 5' ends of mRNAs and even in
XX those cases where longer cDNA sequences have been obtained, the full 5'
XX UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
XX ends and can therefore be used to obtain full length cDNAs and genomic
XX DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
XX chromosome mapping procedures. They are used to obtain upstream
XX regulatory sequences and to design expression and secretion vectors.
XX
SQ Sequence 105 AA;

```

```

Query Match 33.8%; Score 190; DB 21; Length 105;
Best Local Similarity 36.2%; Pred. No. 2; Le-16;
Matches 46; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

```

```

OY 1 MVOIITDNEKFTLTAAAGHKLAVQSSKRCGPKRMFPVPHLEA----- 47
DB 1 mvkqlesktafgealdaagdklvvdfsatwcpckmkpfhfsisekysnvlflvvd 60
OY 48 -----TCHIKTIPTFGMFKRSOKVTLFSRIKRIICYSRSGFMSNLIFFCGADAKLE 100
DB 61 dcgdvasacevkcmptqfkkqkv-----etsgankekle 98
OY 101 AKTOELM 107
DB 99 atlnelv 105

```

```

RESULT 7
AAE05136

```

```

ID ID AAE05136 standard; Protein: 210 AA.
XX
AC AAE05136;
XX
DT 12-SEP-2001 (first entry)
XX
DE Human T cell-derived colony stimulating factor (TC-CSF).
XX
KW T cell-derived colony stimulating factor; TC-CSF; immunostimulant;
KW acquired immune deficiency syndrome; AIDS; cancer; gene therapy;
KW haematopoietic growth factor; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 18..28
FT /label= signal_peptide
FT Protein 29..210
FT /label= Mature_human-TC-CSF-protein
FT Misc-difference 181
FT /note= "Encoded by TAA"
FT Misc-difference 188
FT /note= "Encoded by TAA"
FT Misc-difference 210
FT /note= "Encoded by TAA"
XX
XX US6254861-B1.
XX
XX 03-JUL-2001.
XX
XX 24-JAN-1994; 94US-0180371.
XX
XX 23-MAY-1989; 89US-0356006.
XX 12-JUL-1991; 91US-0729135.
XX 19-AUG-1991; 91US-0747784.
XX 01-NOV-1991; 91US-0788115.
XX 01-JUL-1992; 92US-0906866.
XX
XX (CHOUDHURY) CHOUDHURY C.
XX
XX Choudhury C;
XX
XX WPI: 2001-432041/46.
XX N-PSDB: AAD09895.
XX
XX New T cell-derived colony stimulating factor for treating
XX immune-compromised patients, including acquired immunodeficiency
XX syndrome patients and certain types of cancer patients -
XX
XX Claim 1; Column 49-52; 38pp; English.
XX
XX The invention relates to T-cell derived colony stimulating factors
XX (TC-CSF) and nucleic acid molecules encoding them. TC-CSF which
XX is a novel haematopoietic growth factor is useful for treating
XX immune-compromised patients, including acquired immune deficiency
XX syndrome (AIDS) patients and certain types of cancer patients.
XX TC-CSF DNA is used in gene therapy. The present sequence is human
XX TC-CSF.
XX
SQ Sequence 210 AA;

```

```

Query Match 33.8%; Score 190; DB 22; Length 210;
Best Local Similarity 36.2%; Pred. No. 5e-16;
Matches 46; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

```

```

OY 1 MVOIITDNEKFTLTAAAGHKLAVQSSKRCGPKRMFPVPHLEA----- 47
DB 76 mvkqlesktafgealdaagdklvvdfsatwcpckmkpfhfsisekysnvlflvvd 135
OY 48 -----TCHIKTIPTFGMFKRSOKVTLFSRIKRIICYSRSGFMSNLIFFCGADAKLE 100
DB 136 dcgdvasacevkcmptqfkkqkv-----etsgankekle 173

```


[illegible]

XX	
PN	JF07079780-A.
XX	
PD	28-MAR-1995.
XX	
PF	20-SEP-1993; 93JP-0233361.
XX	
PR	20-SEP-1993; 93JP-0233361.
XX	
PA	(AJIN) AJINOMOTO KK.
DR	WPI; 1995-157850/21.
DR	N-PADB; AAC87817.
XX	
PT	Prepn. of recombinant human ADF by direct expression in E. coli -
PT	for treating inflammation and radiation sickness caused by
PT	commercially produced free radical(s)
XX	
Claim 2;	Page 7-8; 11pp; Japanese.
XX	
CC	The DNA sequence encoding human ADF was inserted into a plasmid which
CC	was then used to transform E.coli. The E.coli was then cultured and
CC	the recombinantly produced ADF harvested. Human ADF can be used as a
CC	treating agent for inflammation and radiation sickness.
XX	
XQ	Sequence 104 AA:

Query Match	32.9%;	Score 185;	DB 16;	Length 104;
Best Local Similarity	35.7%;	Pred. No. 8.9e-16;		
Matches	45;	Conservative	11;	Mismatches 26; Indels 42; Gaps
QY	2	VOIKDITNEKTEFLTAAGHKLAIVQFSSKRCGCPKRPVPFHLEA-----	47	
Db	1	vkgesktafgealdaagklvvdwdfsatwcpckmkpffhsalsekysnviflevdvd	60	
QY	48	-----TCHKTLPTQOMPFKSKQVLLFSRIKRIICQYSGSMNLIPFCGADAKKLEA	101	
Db	61	cqvdsesecevkcmlptqfrikkgkv-----efsgankexlea	98	
QY	102	KIQELM 107		
Db	99	tlnelv 104		
RESULT 12				
AAP94809				
ID	AAP94809	standard;	protein;	172 AA.
XX	AAP94809;			
XX	25-JUN-1990	(first entry)		
XX	Human Interleukin 2:human adult T cell leukaemia derived factor fusion			
XX	protein.			
XX	Human adult T cell leukaemia derived factor; hADF: cancer;			
XX	immunodeficiency disease; Interleukin 2; fusion protein.			
XX	OS			
XX	Homo sapiens.			
XX	Key	Location/Qualifiers		
XX	Region	1..70		
XX	FT	/label= hIL-2		
XX	FT	71..172		
XX	FT	/label= hADF		
XX	EP299206-A.			
XX	18-JAN-1989.			
XX	PD			
XX	10-JUN-1988;	88EP-0109311.		
XX	PF			
XX	XX			

	RK	12-JUN-1987;	87JP-0146348.
XX	PA	(AJIN) AJINOMOTO KK.	
XX	PI	Yodo J , Tagaya Y , Maeda M , Matsui H , Kondo N , Hamuro J ;	
XX	DR	WPI; 1989-016762/03.	
XX	PT	Recombinant human adult T cell leukaemia derived factor polypeptide -	
XX	PR	used for treating cancer, immunodeficiency disease etc.	
XX	PS	Disclosure; : 24pp; English.	
CC	CC	The fusion protein is cleaved withclostripaininto give ADF protein.	
CC	CC	The polypeptide causes differentiation and induces growth of lymphocytes	
CC	CC	and fibroblasts.	
CC	CC	See also AAP92141.	
SO	Sequence	172 AA;	
OY	Query Match	32.7%; Score 184; DB 10; Length 172;	
Db	Best Local Similarity	36.0%; Pred.No.2.3e-15;	
	Matches	45; Conservative 10; Mismatches 28; Indels 42; Gaps 2	
OY	2 VQIINDTEFEKTFLTAAGHKLAVVPFSSKRGCPCRMPEVHELAE-----	47	
Db	70 vqlesktaifgealdaagdkivvdlsatwcpckmlypfhslekysnvlflewdvdd	129	
OY	48 -----TCIIKTIPTPFOMEKKSQKVTLFSRIKIICCVSGMSNLIPEFCGADAKLEA	101	
Db	130 cgdvaseevckmplfgfftkgkgvg-----efsganveklea	167	
OY	102 KTQEL 106		
Db	168 tlneel 172		
RESULT 13	AAR31931		
ID	AAR31931 standard; Protein; 209 AA.		
XX	AAR31931;		
AC	01-JUN-1993 (first entry)		
DT	Human TC-CSF.		
XX	DE		
KM	Clone; human; T cell; colony stimulating factor; TC-CSF; AIDS;		
KW	cancer; restriction site; ligand; immunoassay; hybridisation assay;		
RV	primer.		
OS	Homo sapiens.		
NN	WO9300923-A.		
PD	21-JAN-1993.		
PF	07-JUL-1992; 92MO-US05707.		
XX	12-JUL-1991; 91US-0729135.		
PR	19-AUG-1991; 91US-0747784.		
PR	01-NOV-1991; 91US-0788115.		
PA	(CHOU/) CHOUDHURY C.		
XX	Choudhury C;		
XX	WPI; 1993-045234/05.		
DR	N-PSDB; AAQ35299.		
XX	New T-lymphocyte-derived haematopoietic growth factor (TC-CSF) -		
PT	for diagnosing deficiency or excess of TC-CSF, stimulating		

PT	haematopoiesis and restoring immune function
XX	
PS	Disclosure; Page 87-88; 107pp; English.
CC	
CC	This sequence is encoded by a partial human T cell colony stimulating factor (TC-CSF) clone. Amino acids represented by X in this sequence have been identified as unknown residues. They represent mutations or sequencing errors. TC-CSF polypeptides may be used in pharmaceutical compositions for diagnosis and treatment of immune-compromised patients, including AIDS patients and certain types of cancer patients. TC-CSF and TC-CSF ligands can be used in immunoassay. The TC-CSF coding sequence, or portions of it, may be used in hybridisation assay or as primers.
CC	
CC	
SQ	Sequence 209 AA:
Query Match	32.7%; Score 184; DB 14; Length 209;
Best Local Similarity	35.4%; Pred. No. 2,9e-15;
Matches 45; Conservative 11; Mismatches 29; Indels 42; Gaps 2	
Dy	1 MWQIIKIDNEFKFTPLTAGHKLAVYQFSSKRCGPCRMPVPEHLAE-----47 : : : : : 76 mvkgstsktafgdaagdklvvdtsfswcgpckmkpfhnsjseksynvlfleavd 135
Dy	48 -----TCHIKITIPFPQMFKKSQKVTLFSRIKRIRIICCYRSGFMSNLIFFCGADAKKE 100 : : 136 dcqgdvasceevkcmpllqgfkkgykv-----efsgankekke 173
Dy	101 AKTOELM 107
Dy	: :
Dy	174 atinelv 180
RESULT 14	
AAR42820	AAR42820 standard; Protein; 104 AA.
AC	AAR42820;
XX	
DT	04-MAY-1994 (first entry)
XX	
DE	RECFE 104.
XX	
KW	Eosinophil cytotoxicity enhancing factor; ECEF; stimulation; diagnosis; therapy; thioredoxin.
OS	Homo sapiens.
FN	W09320107-A.
PD	14-OCT-1993.
PF	06-APR-1993; 93WO-US03310.
PR	06-APR-1992; 92US-0862832.
PR	06-JUL-1992; 92US-0906842.
PA	(BGHM) BRIGHAM & WOMENS HOSPITAL.
PI	Balciewicz-sablinska MK, Silberstein DS;
DR	WI; 1993-336837/42.
DR	N-PSDB; AAO49403.
PT	Shortened Eosinophil Cytotoxicity-Enhancing Factor proteins - having Eosinophil-stimulating activity, for use in diagnosis and therapy
PS	Example 3: Page 65; 96pp; English.
CC	A 14 kDa molecule designated various names (Eosinophil Cytotoxicity-Enhancing Factor (ECEF); thioredoxin and Adr) has

CC been found to support the growth of an EBV transformed B
 CC lymphocyte cell line in the manner of IL-1, to induce the
 CC expression of IL-2 receptors in a large granular lymphocyte cell
 CC line from an ATL patient in the manner of IL-1 and to mediate the
 CC growth inhibitory properties of interferon gamma. Recombinant ECFE
 CC (rECFE) 104 is the full length 104 amino acid polypeptide encoded
 CC by this molecule and was produced for comparison studies against
 CC shortened versions of the polypeptide. Some shortened ECFE's are
 CC useful because they retain eosinophil stimulating activity but lack
 CC proinflammatory dithiol reductase activity. In particular they can
 CC be used for killing tumour cells.

XX Sequence 104 AA;

SO Query Match 32.6%; Score 183; DB 14; Length 104;
 Best Local Similarity 35.8%; Pred. No. 1,6e-15;
 Matches 44; Conservative 10; Mismatches 27; Indels 42; Gaps 2;

OY 5 IKDNEFKFTLTAAGHKLAIVQSSKRCGCKRMFPVFEHLAE----- 47
 Db 4 Iesktfgealdaagdklivvdfsatwcgckmnpffhsisekysnvlflevdvdcdg 63
 OY 48 ---TCHIKTIPFQMFKKSKQVTLFSRKRIICCRSGFMSNLIFFCGADAKKLEAKTQ 104
 Db 64 vasecevkcmptqfifkxgqkvg-----etsgankekleatin 101
 OY 105 ELM 107
 Db 102 elv 104

RESULT 15

AAM69567 ID AAM69567 standard; protein; 105 AA.

AC AAM69567;

DT 15-OCT-1998 (first entry)

DE Human thioredoxin mutant C62S/C69S/C73S.

KM Human; thioredoxin; mutant; variant; TRX; AP-1; transcriptional activity;

KW Ref-1.

OS Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 62 /note= "Cys in wild type"

FT Misc-difference 69 /note= "Cys in wild type"

FT Misc-difference 73 /note= "Cys in wild type"

FT

PN EP853088-A2.

PD 15-JUL-1998.

PF 14-JAN-1998; 98EP-0100535.

PR 14-JAN-1997; 97JP-0004489.

PA (ORIT) ORIENTAL YEAST CO LTD.

PA (YODO/) YODOI J.

PI Yodoi J;

DR WPI; 1998-364621/32.

PT New thioredoxin variants with amino acid substitutions for cysteine
 PT except at active centre for increasing stability in non-reducing

PT conditions - used with Ref-1 to enhance AP-1 transcriptional
 PT activity

PS Example 3; Page -; 14pp; English.

CC The present sequence represents a human thioredoxin (TRX) variant
 CC derived from the wild-type sequence (obtained from Genbank), as stated
 CC in the specification. TRX variants can be made stable under non-reducing
 CC conditions by substituting at least one or all Cys residues by other
 CC amino acid residues, except that Cys residues in the active centre of
 CC TRX remain unmodified. Also described in the present invention is a
 CC factor for enhancing the transcriptional activity of AP-1 which
 CC comprises the TRX variant and Ref-1 associated with each other via an
 CC S-S bond. TRX variants can be used with Ref-1 to enhance AP-1
 CC transcriptional activity. TRX variants do not form multimers and are
 CC stable under non-reducing conditions.

SO Sequence 105 AA;

Query Match 32.0%; Score 180; DB 19; Length 105;
 Best Local Similarity 35.4%; Pred. No. 3.9e-15;
 Matches 45; Conservative 11; Mismatches 29; Indels 42; Gaps 2;

OY 1 MVQIIKDTNEFKFTLTAAGHKLAIVQSSKRCGCKRMFPVFEHLAE----- 47
 Db 1 mvkIesktfgealdaagdklivvdfsatwcgckmnpffhsisekysnvlflevdvd 60
 OY 48 -----TCHIKTIPFQMFKKSKQVTLFSRKRIICCRSGFMSNLIFFCGADAKKLE 100
 Db 61 dsqdvaseevkstptqfifkxgqkvg-----etsgankekle 98
 OY 101 AKTQELM 107
 Db 99 atinelv 105

Search completed: June 8, 2002, 04:28:04
 Job time: 4048 sec

PD - 6/5/1998

p1 d-2

Mon Jun 24 08:35:20 2002

IntelliGenetics

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file Seq18Inv.res made by tport on Mon 24 Jun 102 8:30:55-PDT

```
Query sequence being compared: US-09-746-783-18' (1-481)
Number of sequences searched: 2
Number of scores above cutoff: 2
```

Results of the initial comparison of US-09-746-783-18' (1-481) with
 File : x54539.seq
 File : x77584.seq

STDEV	SCORE
24	1
21	1
19	1
16	1
13	1
11	1
8	1
5	1
3	1
0	1

PARAMETERS

	Unary	K-tuple
Similarity matrix	1	4
Mismatch penalty	1	3G
Gap penalty	5.00	Joining penalty
Gap size penalty	0.33	Window size
Cutoff score	1	
Randomization group	0	

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	24	25	0.00

	CPU	Total Elapsed
Times:		

Number of residues:	1384
Number of sequences searched:	2
Number of scores above cutoff:	2

The scores below are sorted by initial score.
Significance is calculated based on initial score

A 100% identical sequence to the query sequence was not found

The list of best scores is:

Seq18.inv.res

SEE ID NO: 18
INTELLIGENCE, Fgtd08
AC NO: X77584, X54539

Page 1

Sequence Name	Description	Length	Int. Score	Opt. Score	Sig. Frame
---------------	-------------	--------	------------	------------	------------

1. x77584	TOIG of: x77584	check: 4885	583	24	166	0.00	0
2. x54539	TOIG of: x54539	check: 1368	801	24	87	0.00	0

1. US-09-746-783-18' (1-481)
x77584 TOIG of: x77584 check: 4845 from: 1 to: 583

Initial Score	=	24	Optimized Score	=	166	Significance	=	0.00
Residue Identity	=	38%	Matches	=	187	Mismatches	=	294
Gaps	=	10	Conservative Substitutions				=	0

CGAGCTTGAAGCTCTGTTGGTCTTTCATTCATTCCATGGTCTTACAGCCGCTGTGCAGACTCAG

10 20 30 40 50 60 70

50 60 70 80 90 100 110
AAGTATATTTGGCCTTGSAGATC-AGCTTACATTAATCTCTAGTCTTGGCTTCCAAATTTTTACACATCGGCG
CAGCCAGATGGTGAAGCAGATCAGAGCAAGACTGCTTTTCAGGAAGCCTTGGACGCTGCAGGTGATTAAC
80 90 100 110 120 130 140

[illegible]

190
GAGATAGGGTACCTTCCTGGCTTTCTTGAA--CATCTGAAGAATGGGTATTGTTTGCATGACCAATTT
| | | | |
200
| | | | |
210
| | | | |
220
CTGAAGAATTCCTCCACCGCAGCATTCCTCGAAGTGCATGTGCAGACTGCACAGATGCTTCACAGCTGC
230
240
250
260
270
280

[illegible][illegible]

410 420 430 440 450 460 470
--TTACACCAGGGAAGTGTGATGAAATCCCTGTGTGGTTACTGGATCACGTAGCTGTCTCCTATT
| | | | | | | | | | | | | | | | | |
ATTTAACCTTGTAATTTTTTTAATTACAAAAATATAAATATAGAAGACAATAAACCCAGTTGCCATCTGC

440 450 460 470 480 490 500

480
TTACGATCC
| | | |
TGCAATAAACATTATGCTAACACTTTTAAACCGTCGATGCTGAATAGCPTTCA
510 X 520 530 540 550 560

2. US-09-746-783-18' (1-481)
x54539 TOIG of: x54539 check: 1368 from: 1 to: 801

Initial Score	=	24	Optimized Score	=	87	Significance	=	0.00
Residue Identity	=	31%	Matches	=	92	Mismatches	=	196
Gaps	=	1	Conservative Substitutions	=			=	0

[illegible]

```
CGAGCTGGGCTTGGGTTACAGAGGCCGAAGCCCTGAGCTCCGCCACAGCTTGGCCAGCCGCCCGA
470      480      490      500      510 X      520      530
      30      40      50      60      70      80      90
TTTTTTTTGCTTTTCAAAATGTACACAGTATTTTGGCTTGGAGATCAGCTTACATTAAATCTTGAGCT
      100      110      120      130      140      150      160
TGCTTCCAAATTTTATGATCGGCTCCACAAAATCAAAAATCAGGTTGCTCATGATCACTTCTATAC
      170      180      190      200      210      220      230
AGCAAAATTAATCTTTGATCTTGAAGAAATAGG-GTTACCTTCTGCTTTTCTTGAACATCTGAATGTGGT
      240      250      260      270      280      290      300
ATTGTTTGAATGACACAGTTTCAGCCAGCTCAATGCAAAACAGGAAACATCTTTTGCAGGGACCAACCGT
      310      320      330
TTGGAAGAAATTTGAACCACTGGAGTTT
```


Use for 101

Not Cited
Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2002, 02:13:55 ; Search time 1874.6 seconds
(without alignments)
5369.300 Million cell updates/sec

Title: US-09-746-783-18

Perfect score: 481

Sequence: 1 GCATACCTGTAATAATAGCA.....AAAAAAAAAAAAAAAAAAAA 481

Scoring table:

IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

GenBank: 1: gb_ba: 2: gb_hgt: 3: gb_in: 4: gb_om: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_sts: 12: gb_sy: 13: gb_un: 14: gb_vl: 15: em_ba: 16: em_fun: 17: em_hum: 18: em_in: 19: em_mu: 20: em_om: 21: em_or: 22: em_ov: 23: em_pat: 24: em_ph: 25: em_pi: 26: em_ro: 27: em_sts: 28: em_un: 29: em_vl: 30: em_hgt_hum: 31: em_hgt_inv: 32: em_hgt_other: 33: em_hgt_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Match Length DB ID Description

C	1	114	✓23.7	194835	9	AL158158	✓	AL158158	Human DNA
	2	97	✓20.2	214852	9	CNS05TCC	✓	AL355888	Human chr
	3	79.4	16.5	255	9	AF065241		AF065241	Homo sapi
	4	63.8	13.3	350	4	AF104105		AF104105	Bos tauru
	5	62.2	12.9	532	4	OATHORD		Z25864	O.aries mRN
	6	61.6	12.8	382	6	AX321985		AX321985	Sequence
	7	61.4	12.6	371	4	AF382821		AF382821	Sus scrofa
	8	60.6	12.6	583	6	E01915		E01915	CDNA encodi
	9	60.6	12.6	583	9	HSATLRED		X77584	H. sapiens m
	10	60	12.5	529	9	BC003377		BC003377	Homo sapi
	11	60	12.5	630	9	AR160057		AR160057	Sequence
	12	59.8	12.4	371	9	AF313911		AF313911	Homo sapi
	13	59.6	12.4	477	9	HUMY146D09		AY004872	Homo sapi
	14	58.8	12.1	439	9	HUMY146D09		AY004872	Homo sapi
	15	58.4	12.1	501	9	HUMY146D09		AY004872	Homo sapi
	16	57.8	12.0	364	9	AF353204		AF353204	Human thior
	17	57.8	12.0	591	5	AF321769		AF321769	Ophiophag
	18	57.2	11.9	421	9	AF276919		AF276919	Homo sapi
	19	56.8	11.8	506	9	MACYRDA		M84643	Macaca mula
	20	56.8	11.8	630	6	AR160060		AR160060	Sequence
	21	56	11.6	335	6	E05024		E05024	Human ADF
	22	56	11.6	315	6	E05929		E05929	DNA encodin
	23	56	11.6	315	6	E08965		E08965	DNA encodin
	24	55.2	11.5	562	9	AF146023		AF146023	Homo sapi
	25	55.2	11.5	661	5	CHKTHD		J03882	Chicken th
	26	55.2	11.5	117930	9	AC073271		AC073271	Homo sapi
	27	55.2	11.5	161081	2	AC022997		AC022997	Homo sapi
	28	55.2	11.5	163319	2	AC108512		AC108512	Homo sapi
	29	55	11.4	1764	9	HSAA20593		AJ420593	Homo sapi
	30	53.6	11.1	535	4	AB022431		AB022431	Equus cab
	31	53	11.0	312	6	E06575		E06575	CDNA encodi
	32	53	11.0	312	6	E08964		E08964	DNA encodin
	33	51.2	10.6	1210	9	AF146024		AF146024	Homo sapi
	34	51.2	10.6	177750	9	AC005874		AC005874	clbL175-
	35	51.2	10.6	17864	9	AF134471		AF134471	Homo sapi
	36	51.2	10.6	178964	2	AC012468		AC012468	Homo sapi
	37	51	10.6	4938	3	AY069493		AY069493	Drosophil
	38	50.2	10.4	534	10	BC010756		BC010756	Mus muscu
	39	50.2	10.4	538	10	MMTRX		X77585	M. musculus
	40	50	10.4	362	6	A84204		A84204	Sequence 29
	41	50	10.4	349980	6	AX344558		AX344558	Sequence
	42	49.8	10.4	311	3	AF159977		AF159977	Buthus ma
	43	49.8	10.4	1577	9	BC011971		BC011971	Homo sapi
	44	49.8	10.4	63374	2	AC102559		AC102559	Mus muscu
	45	49.4	10.3	5383	5	AB027537		AB027537	Xenopus 1

ALIGNMENTS

RESULT 1
AL158158/c
LOCUS
DEFINITION
Human DNA sequence from clone Rp11-427/L11 on chromosome 9q31.2-32,
complete sequence.
ACCESSION
AL158158
VERSION
AL158158.14 GI:13559997
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
On Apr 6, 2001 this sequence version replaced gi:12733508.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission


```

misc_feature complement(23067..23458)
              /note="match: GSS: Em:B43385"
misc_feature complement(23074..23458)
              /note="match: GSS: Em:B43132"
repeat_region repeat_region
              /note="MER5A repeat: matches 17..180 of consensus"
misc_feature 24008..24365
              /note="match: GSS: Em:AQ088039"
repeat_region repeat_region
              /note="MIR repeat: matches 105..202 of consensus"
repeat_region repeat_region
              /note="LIMB8 repeat: matches 6145..6287 of consensus"
repeat_region repeat_region
              /note="MER50 repeat: matches 1..436 of consensus"
repeat_region repeat_region
              /note="MER50 repeat: matches 640..734 of consensus"
repeat_region repeat_region
              /note="LIME1 repeat: matches 5212..5366 of consensus"
repeat_region repeat_region
              /note="ALUSp/q repeat: matches 174..260 of consensus"
repeat_region repeat_region
              /note="LIME1 repeat: matches 5347..6036 of consensus"
repeat_region repeat_region
              /note="LIMB6 repeat: matches 4066..4176 of consensus"
repeat_region repeat_region
              /note="ALUSx repeat: matches 1..310 of consensus"
repeat_region repeat_region
              /note="LIMB6 repeat: matches 4176..4455 of consensus"
repeat_region repeat_region
              /note="ALUB repeat: matches 3..132 of consensus"
repeat_region repeat_region
              /note="ALUS30 repeat: matches 1..297 of consensus"
repeat_region repeat_region
              /note="ALUS31 repeat: matches 1..28677"

```

```

Query Match 23.7%; Score 114; DB 9; Length 194835;
Best Local Similarity 92.3%; Pred. No. 5.8e-16;
Matches 120; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 330 TTCATGAGCAACGATTTTGAAGTTTGTGAGCCGATGCTAAATAATTGAGCCCAAG 389
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67911 TGCATTGTGACAGATTTTGAAGTTTGTGAGCCGATGCTAAATAATTGAGCCCAAG 67852

QY 390 ACTCAGAAATTAATGATGATCTCCAGCAAAATATACATTTGACATTTGAAAAAG 449
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67851 ACTCAGAAATTAATGATGATCTCCAGCAAAATATACATTTGACATTTGAAAAAG 67792

QY 450 CAAAAAAA 459
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67791 CAAAGACAAA 67782

```

```

RESULT 2
CNS05TC2 214852 bp DNA linear PRI 29-MAY-2001
LOCUS DEFINITION Human chromosome 14 DNA sequence BAC R-322L17 of library RCI-11
ACCESSION AL355888
VERSION AL355888.3 GI:14272176
KEYWORDS HMG; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 214852)
Hellig, R., Pett, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,
Brotier, P., Cattolico, L., Barbe, V., Pelletier, E., Arliguenave, F.,
Levy, M., Eckenberg, R., Bruls, T., deBertandis, V., Crnaud, C.,
Gyapay, G., Saurin, W. and Weissenbach, J.
JOURNAL Unpublished
REFERENCE Sequencing of the human chromosome 14
AUTHORS 2 (bases 1 to 214852)
TITLE Genoscope.
TITLE Direct Submission

```

JOURNAL Submitted (29-MAY-2001) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 Web : www.genoscope.cns.fr)
 On May 31, 2001 this sequence version replaced gi:9886699.
 ----- Genome Center
 Center: Genoscope / Centre National de Sequencage
 Center code: GS
 Web site: http://www.genoscope.cns.fr/
 Contact: SeqRef@genoscope.cns.fr

The following BAC sequence is oriented from the T7 to the SP6 end.
 Upstream BAC (overlapping the T7 end) : R-305B6
 Downstream BAC (overlapping the SP6 end) : R-99L13 (AC-AL161752)
 ----- Summary Statistics
 Assembly program: Phrap; Statistics
 Quality coverage: 6.02x in Q20 bases; sum-of-contigs

 Overall quality chart :
 Range : bases

```

0 - 9 : 2
10 - 19 : 29
20 - 29 : 127
30 - 39 : 494
40 - 49 : 4696
50 - 59 : 14835
60 - 69 : 19064
70 - 79 : 37628
80 - 89 : 65006
90 - 99 : 72971

```

Percentage of bases with a quality value >= 40 : 99 %.

FEATURES
 source Location/Qualifiers

```

1..214852
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="14"
  /clone="R-322L17"
  /clone_1b="RCI-11"
  181957..182185
  /note="matchling EMBL:G02797
  RhdB:RH53844
  RhdB:RH3727
  dbSTS:STS14441
  Identified using the e-PCR software (G. Schuler)"
BASE COUNT 74248 a 38130 c 37571 g 64903 t
ORIGIN

```

```

Query Match 20.2%; Score 97; DB 9; Length 214852;
Best Local Similarity 87.6%; Pred. No. 3.7e-12;
Matches 106; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```

```

QY 339 AACCTGATTTTGAAGTTTGTGAGCCGATGCTAAATAATTGAGCCCAAGACTCAGAA 398
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 91231 AACGAGACATTTAGATTGTTGGGGCGTGAATAAAATTTGAGCCCAAGACTCAGAA 91290

QY 399 TTAATGTAAGTGAATTCGCAAGCAAAATACACTTTGGACATTTGAAAAGCAAAAAA 458
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 91291 TTAATGCAAGCTGATCTCCAGCAAAATATATCTTGACATTTCAAAAAGCAAAAAA 91350

QY 459 A 459
    |
Db 91351 A 91351

```

```

RESULT 3
LOCUS AF065241 255 bp mRNA linear PRI 27-MAY-1998
DEFINITION Homo sapiens thioresdoxin delta 3 (TXN delta 3) mRNA, partial cds.
ACCESSION AF065241
VERSION AF065241.1 GI:3153858

```

FEATURES	source	location/qualifiers
gene	CDS	<1..255 /gene="TXN delta 3" /note="TXN delta 3" /note="alternative splice variant" /codon_start=1 /product="thioredoxin delta 3" /protein_id="AAC17430.1" /db_xref="GI:3153859" /translation="VKQESTKTAQELADLAAGDKLVVYDPSATWCGPCKMKIPFFHDV ASCEVEMCPFFPKFGKQKGFSGNKEXLETINELY" ASCEVEMCPFFPKFGKQKGFSGNKEXLETINELY"
BASE COUNT	74 a 45 c 66 g 70 t	
ORIGIN		
Query Match	16.5%	Score 79.4; DB: 9; Length 255;
Best Local Similarity	60.4%	Pred. No. 5.1e-08;
Matches 131;	Conservative 0;	Mismatches 86; Indels 0; Gaps 0;
QY	87 GTACAGATTATTAAAGACAGCATGATGATTTAAACATTTTTCAGACGTCGCGACACAAA	146
Db	1 GTGAAGCAGATCGAGAGCAAGACGCTTTTCAGGAAGCCTTGAGCGTCGAGGTGATAA	60
QY	147 CTCGACGTGTCATATTTTCCTTCGAAAGGCTGTGTCCTCCCTGCAAAAGATGTTCTCTT	206
Db	61 CTGTGAGTAGTTCATCTTCTTCAGCCAGCTGTGGCTTCGAAATATATCAAGCTTTC	120
QY	207 TTCATGAGCTGGCGCAACTTGTTCACATCAAAACATATCCACATTTTCAGATGTTCAAG	266
Db	121 TTTGATGATGTTGCTTCAGAGGTGATGATTCATATGCAATGCAATGCAATGTTTAAAG	180
QY	267 AAAAGCCAGAAGGTACCCCTATTTCTCAAGAATCAAAA	303
Db	181 AAGGACAAAAGGTGGTGAATTTTCTGAGGCCAATTA	217
RESULT 4		
AF104105	AF104105	550 bp mRNA linear MAM 28-OCT-1999
DEFINITION	Bos taurus thioredoxin mRNA, complete cds.	
ACCESSION	AF104105	
VERSION	AF104105.1	GI:3978528
KEYWORDS		
SOURCE	cow.	
ORGANISM	Bos taurus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	Wang, Y., Wang, Y. G., Zhang, Y., Yuan, Y., and Ma, D.	
TITLE	An alternative splice variant of human thioredoxin	
JOURNAL	Chin. Sci. Bull. 43, 292-295 (1998)	
REFERENCE	2 (bases 1 to 255)	
AUTHORS	Wang, Y., Wang, Y. G., Zhang, Y., Yuan, Y., and Ma, D.	
TITLE	Direct Submission	
JOURNAL	Submitted (13-MAY-1998) Molecular Immunology, Basic Medical School, Beijing Medical University, Xueruan Road 38, Beijing 100083, P. R. China	

FEATURES	source	location/Qualifiers
REFERENCE	20112351	
AUTHORS	Gotoh,S., Terashima,H., Yagi,K. and Mizoguchi,T.	
TITLE	Submitted Submission	
JOURNAL	Direct Submission	
	Submitted (02-NOV-1998) Bio-functional Molecular Chemistry, Faculty of Pharmaceutical Sciences, Osaka University, Yamadaoka 1-6, Suita, Osaka 565, Japan	
source	location/Qualifiers	
	1..550	
	/organism="Bos taurus"	
	/db_xref="taxon:9913"	
CDS	76..393	
	/codon_start=1	
	/product="thioredoxin"	
	/protein_id="AAC83380.1"	
	/db_xref="GI:3978529"	
	/translation="MVQIESKVFQFALMSAGEKLVYVDFSATWGPCMKIRPFHS	
	LSERYSNVFLEVEDDCCDVAAECVECKMPTTFQFFKGGKGVGSEANKREKLEATIN	
	ELI"	
BASE COUNT	167 a 107 c 116 g 160 t	
ORIGIN		
Query Match	13.3%; Score 63.8; DB 4; Length 550;	
Best Local Similarity	53.0%; Pred. No. 0.00015;	
Matches 184; Conservative	0; Mismatches 157; Indels 6; Gaps 2;	
QY	71	CCCTGCTGTATTCATGATGATGATATTTAAAGACAGCAATGATTTAAACATTTTGGAC 130
DB	63	CCGACCTGCGCAAGATNGTGGAAACAGATGAGACAGCAAGATGCTTTTCAGAAAGCCTTGAA 122
QY	131	AGCTGCGGAGACAAATCGCATGCGATGCTTCAATTTTCTTGCAGAACGGTGTGCTCCCTGCA 190
DB	123	CAGTGCAGAGAGAAAGAACTCGTAGTAGTGCAGCTTCCACCCACGAGTGGTGGCCCTTGCAA 182
QY	191	AAGGATGTTCTCTGTTTCCATGAGCTGGCGGAAACCTGTGCATCAAAACATATCCAC 250
DB	183	AATGATCAAGCCTTCTCTTCTTCAATCTCTCTGGAAGAAGTATTCACAGCTGGTCTTGGA 242
QY	251	ATTTCAGATGTTTCAGAAAGACCAAGAGTAACCTATTTCTCAAGAAATCAAAAGATTAAT 310
DB	243	AGTAGATGTG--GATGACTGTCAAGAGTGTGTGCAAGTAGTGAAGTCAATATCATGATGCC 299
QY	311	TTGCGTATATAGAGTGGATTCATGAGCAACCTGATTTTGTAGTTTGTGGAGCGGATGC 370
DB	300	AACCTTCCAGTTTTTTAAA--AAGGGACAGAAAGTGGGTAAATTTCTGGAGGTAATAA 356
QY	371	TAAATAATTGGAGCCAGACTCAAGAATTAATGTAATGATGATCTCC 417
DB	357	AGAAAACTTGAGCGCACCATTAATGATTAATCATCATGATTTTC 403
RESULT	5	
OATHIORD	OATHIORD	532 bp mRNA linear MAM 31-MAY-1995
LOCUS	O.aries mRNA for thioredoxin.	
DEFINITION	Z25864	
ACCESSION	Z25864.1	GI:397946
VERSION		
KEYWORDS	thioredoxin.	
SOURCE	sheep.	
ORGANISM	Ovis aries	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;	
	Bovidae; Caprine; Ovis.	
REFERENCE	1 (bases 1 to 532)	
AUTHORS	Droogmans,L., Cleuter,Y., Wollman,E.E., Kettmann,R. and Burny,A.	
TITLE	Nucleotide sequence of ovine thioredoxin cDNA	
JOURNAL	DNA Seq. 4 (4), 277-279 (1994)	
MEDLINE	95078463	
REFERENCE	2 (bases 1 to 532)	
AUTHORS	Droogmans,L.	
TITLE	Direct Submission	
JOURNAL	Submitted (01-SEP-1993) Droogmans L., Universite Libre de	

source	1..583	/organism="Homo sapiens"
		/db_xref="taxon:9606"
BASE COUNT	178 a 111 c 120 g 174 t	
ORIGIN		
Query Match	12.6%;	Score 60.6; DB 6; Length 583;
Best Local Similarity	52.1%;	Pred. No. 0.00079;
Matches 185; Conservative	0;	Mismatches 164; Indels 6; Gaps 2
OY	62	TCAGCATTCCCTGGTGTATCATCATGTACAGATTATTAAGACAGCAATGAATTTAAAC 121
Db	59	TCGTGAGACTGCAGACGCCAAGATGTGTAGACGATCGAGACAGACCTGTTTCAGGA 118
OY	122	ATTTTGACAGTCGCCGACACAACTGCAGCTGGTTCATTTCTTGGAAAGGCTGG 181
Db	119	AGCCTTGACAGCTGCAGGTGATTAACCTGTAGTACTGCTCTAGCCACGCTGGTGG 178
OY	182	TCCCTGCAAAAGATGTTTCTCTGTTTCCATGAGCTGGCTGAACCTGTACATCAAAAC 241
Db	179	GCCTTGCAAAATGATCAACACCTTTCTTCATTCCTCTCGAAAGATATCCACAGTGT 238
OY	242	AATACCACATTTGAGTGTTCAGAAAAAGCCAGAAAGGTAACCCATATCTCAAGATCA 301
Db	239	ATTCTTTGAAAGTAGATGTG--GAGGACTGTGAGATGTGCTTCAAGACTGTGAAGTCA 295
OY	302	AAGAATAATTTGCTTTTATAGAGTGCATTCATGAGCAACCTGATTTTGGACTTTTGG 361
Db	296	ATGCAATGCCACATTCACGTTTTTTAAG--AAGGACAAAGGTGGGTGAATTTTCTGG 352
OY	362	AGCCGATCTTAAAAATTGGAAGCCAGACTCAAGATTAATGTAACTGATCTC 416
Db	353	AGCCATAAGGAAAGCTTGAAAGCCACCATTAATGAATTAATGATATCATCATGTTTC 407
RESULT 9		
HSATLIED		
LOCUS	583 bp	mRNA linear PRI 09-FEB-1994
DEFINITION	H.sapiens mRNA for ATL-derived factor/thioredoxin.	
ACCESSION	X77584	
VERSION	X77584.1	GI:453963
KEYWORDS	thioredoxin.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	Tagaya,Y., Medea,Y., Mitsui,A., Kondo,N., Matsui,H., Hamuro,J., Brown,N., Arai,Ki., Yokota,T., Wakasugi,H. and Yodoi,J.	
TITLE	ATL-derived factor (ADF), an IL-2 receptor/Tac inducer homologous to thioredoxin; possible involvement of dichloro-reduction in the IL-2 receptor induction	
JOURNAL	EMBO J. 8 (3), 757-764 (1989)	
MEDLINE	89251607	
REMARK	Erratum: [[published erratum appears in EMBO J 1994 May 1;13(9):2244]]	
REFERENCE	2 (bases 1 to 583)	
AUTHORS	Tagaya,Y.	
TITLE	Direct Submission	
JOURNAL	Submitted (07-FEB-1994) Y. Tagaya, Division of Allergy Immunology, Dept of Medicine, Cornell Medical Center, 525 East 68th Street, Rm. LC 907, New York, NY 10021, USA	
FEATURES	Location/Qualifiers	
SOURCE	1..583	/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="1"
		/cell_line="ATL-2 (human HTLV-I transformed T cell line)"
		/clone_id="ATL-2/19t11"
		81..398
		/codon_start=1
		/product="ATL-derived factor/thioredoxin"

Probable match to Homo sapiens protein P10599 (PID:g2982059) Human Thioresdoxin (Oxidized With Diamide) Oxidoreductase, Dimer, Thioresdoxin, X-Ray Crystallography, Electron Transport Mol_id: 1; Molecule: Thioresdoxin; Chain: Null

The location of this clone is unknown.

FEATURES
source
Location/Qualifiers
1..477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="142289"
/clone_lib="Soares_placenta_1NHP"
misc_feature
13..327
/note="Probable match to Homo sapiens protein J04026 (PID:g339649)"
misc_feature
13..327
/note="Probable match to Homo sapiens protein P10599 (PID:g2982059)"
BASE COUNT 159 a 84 c 96 g 138 t
ORIGIN

Query Match 12.4%; Score 59.6; DB 9; Length 477;
Best Local Similarity 52.3%; Pred. No. 0.0013;
Matches 181; Conservative 0; Mismatches 159; Indels 6; Gaps 2;

OY 72 CCTGCTATATCATGTATACAGATTTATTAAGACACGATGATTTAAACATTTTGGACA 131
DB 1 CCAGACGCCAAGATGCTGAGACGATCGAGACAGATGCTTTTCAGAAAGCCTTGAC 60
OY 132 GCTGCCGACACAACTCCAGATGTTCAATTTTTCGAAACGGTGGTCCCTGCAAA 191
DB 61 GCTGACGATGATTAATTTAGTGTGCTTCTCAGACGATGCTGGCTTGGCAA 120
OY 192 AGATGTTTCTGTTTTCATGAGTGGCTGAAATTTTCACATCAAAACATATCCACA 251
DB 121 ATGATCAAGCCCTTCTTCTCCCTGCTGAAATATTCACAGTATATCTTGA 180
OY 252 TTTCAGATTTCAAGAAAGCCAGAGTAACTTCTCAGATTCAAAGATTAAT 311
DB 181 GTAGATGTC---GATGACTGTGAGATGTTGCTTCAGAGTGAATCAATGCCA 237
OY 312 TCTGTTATGAGATGATTCATGACCAACCTGATTTTGTGAGTGGACCGATGCT 371
DB 238 ACATTCACATTTTAAAG---AAGGACAAAGAGTGGGTAATTTCTGAGCAATTAAG 294
OY 372 AAAAAATGGAAGCCAGACTCAAGATTAATTAAGTGAATCTCC 417
DB 295 GAAAGCTTGAAGCCACATTAATGATTAATCAATGTTTTC 340

RESULT 14

AT004872 439 bp mRNA linear PRI 21-JUL-2000
LOCUS Homo sapiens thioresdoxin (TXN) mRNA, complete cds.
DEFINITION Homo sapiens thioresdoxin (TXN) mRNA, complete cds.
ACCESSION AY004872
VERSION AY004872.1 GI:9508996
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 439)
AUTHORS Liu, A. and Lou, M.F.
TITLE Cloning, purification and characterization of human lens thioresdoxin
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 439)
AUTHORS Liu, A. and Lou, M.F.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-2000) Veterinary & Biomedical Sciences, University of Nebraska at Lincoln, 120 VBS, East Campus, Lincoln, NE 68583-0905, USA

FEATURES
source
Location/Qualifiers
1..439
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HLE-B3"
/cell_type="epithelial"
/tissue_type="lens"
1..439
/gene="TXN"
CDS
11..328
/gene="TXN"
/note="oxidoreductase"
/product="thioresdoxin"
/protein_id="AAF87085.1"
/db_xref="GI:9508997"
/translation="MYKQIESKTAPEALDAAGDKLVVDFSAITWGPCMKIPKFFHS
LSEKSNVIFLEVDVDDCODVASECEVKCMPTFQFFKKGQKVEFGANKREKLEATIN
ELV"

BASE COUNT 138 a 78 c 95 g 128 t
ORIGIN

Query Match 12.2%; Score 58.8; DB 9; Length 439;
Best Local Similarity 52.7%; Pred. No. 0.002;
Matches 176; Conservative 0; Mismatches 152; Indels 6; Gaps 2;

OY 84 ATGTACAGATTTATTAAGACACGATGATTTAAACATTTTGCAGCTGCCGACAC 143
DB 11 ATGTGACAGCAGATCGAGACAGATGCTTTTCAGAAAGCCTTGACGATGAT 70
OY 144 AATCGACAGTGGTCAATTTTCTGAAACGGTGGTCCGCAAAAGATGTTCT 203
DB 71 AATCTGTACTAGTGTGCTTCTGACGACGCTGGTGGCTTGCATAATGATCAACCT 130
OY 204 GTTTCATGAGCTGGTGGTAACTTGCACATCAAAACATATTCAGATGTC 263
DB 131 TTCTTCATTCCTCTCTGAAAGTATTCACAGTATATTCCTGAAATGATGTC--- 187
OY 264 AAAAAAGCCAGAGATTAACCTATTTCTCAAGATCAAAAGATTAATTTCTGTATAGA 323
DB 188 GATGACTGTGAGATGTTGCTTCAGATGTAAGTCAAAATGACCAATTCACATTT 247
OY 324 AGTGAATTCAGCAACCTGATTTTGTGAGTGGTGGACCGGCTGAATAATGGA 383
DB 248 TTTAAG---AAGGACAAAGAGTGGTGAATTTTCTGAGCCATTAAGAAAGCTTGA 304
OY 384 GCCAGACTCAAGATTAATTAAGTGAATCTCC 417
DB 305 GCCACCATTAATGATTAATGATTAATGATTTTC 338

RESULT 15

HUMTHD 501 bp mRNA linear PRI 23-AUG-1995
LOCUS Human thioresdoxin (TXN) mRNA, complete cds.
DEFINITION Human thioresdoxin (TXN) mRNA, complete cds.
ACCESSION J04026
VERSION J04026.1 GI:339648
KEYWORDS
SOURCE Human cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 501)
AUTHORS Wolman, E.E., d'Aurion, L., Rimsy, L., Shaw, A., Jacquot, J.P., Wingfield, P., Graber, P., Desars, F., Robin, P., Galibert, F., Bertoglio, J., and Fradeliz, D.
TITLE Cloning and expression of a cDNA for human thioresdoxin
JOURNAL J. Biol. Chem. 263 (30), 15506-15512 (1988)
REFERENCE 2 (bases 1 to 501)
AUTHORS Wolman, E.E.
TITLE Direct Submission

JOURNAL

Submitted (24-AUG-1988) E.E. Wollman, Centre National de la
Recherche Scientifique, UA 1156/Institut National de la Sante et de
la Recherche Medicale/Institut Gustave Roussy, Villejuif, France

FEATURES

SOURCE

1..501

/organism="Homo sapiens"

/db_xref="taxon:9606"

/map="3p12-p11"

/cell_line="3B6"

/tissue_type="EBV transformed lymphocytes"

1..501

/gene="TXN"

64..381

/gene="TXN"

/codon_start=1

/product="thioredoxin"

/protein_id="AA74596.1"

/db_xref="GI:33649"

/db_xref="GDB:G00-120-475"

/translation="MKQIESKTAFOEALDPAQDKLVVDEFSATWGPCKINPEFHS
LSEKYSNVIFLEVVDQDQVASECEVKCTPEQFQFKGQKVEFGSANKKELEATIN
ELY"

gene

CDS

BASE COUNT 149 a 100 c 105 g 147 t
ORIGIN 1 bp upstream of EcoRI site.

Query Match

12.1% Score 58.4; DB 9; Length 501;

Best Local Similarity 51.7%; Pred. No. 0.0025;

Matches 184; Conservative 0; Mismatches 166; Indels 6; Gaps 2;

```
QY 62 TCAGACTTCCTGGTGAATCATGCTGACATATTTAAAGACAGATGAATTTAAAC 121
    || || || || || || || || || || || || || || || || || || ||
Db 42 TCGTCAGACTCCACGACGACGAGATGATGAGCAGATCGAGCAAGACTGCTTTCAGGA 101

QY 122 ATTTTGCAGCTCCGACACACAAAACGTCAGTGTCAATTTCTTGAACGGGTGG 181
    || || || || || || || || || || || || || || || || || || ||
Db 102 AGCCTTGAGCGCTGCAGGTGATTAACCTGTAGTACTTCTCAGCAGCTGCTGG 161

QY 182 TCCCTGCAAAAGATGTTCTCTGTTTCATGAGCTGCTGAACCTGTCAATCAAAAC 241
    || || || || || || || || || || || || || || || || || || ||
Db 162 GCCTTGCAAAATGATCAACCTTCTTTCATTCCTCTCTGAAAAGTATCCAAAGTGA 221

QY 242 AATACCCACATTTCAAGATGTCAGAAAAGCCAGAGTAACCTATTCACAGATCAA 301
    || || || || || || || || || || || || || || || || || || ||
Db 222 ATTCCTTGAAATGATGATG--GATGACTGTGACGATGTGCTTCAGAGTGAAGTCAA 278

QY 302 AAGAATAATTTGCTGTATAGAGTGAATGATGAGCAACTGATTTTGAATTTGGTGG 361
    || || || || || || || || || || || || || || || || || || ||
Db 279 ATGACAGCCACATTCACATTTTAAAG--AAGGACAAAAGGTGGGTGAATTTTCTGG 335

QY 362 AGCGGATGCTAAAAAATTTGAGACCAAGACTCAAGAAATTAATGAAGCTGATCTCC 417
    || || || || || || || || || || || || || || || || || || ||
Db 336 AGCCAATTAAGAAAAGCTTGAAGCCACCATTAATGAATTAAGTCAATCATGTGTTTC 391
```

Search completed: June 8, 2002, 03:20:32
Job time: 3997 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2002, 02:15:20 ; Search time 78.88 seconds
(Without alignments)
1497.841 Million cell updates/sec

Title: US-09-746-783-18

Perfect score: 481

Sequence: 1 GCATACCTGTAATAATAGCA.....AAAAAAAAAAAAAAAAAAAA 481

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/prodata/2/lna/5A_COMB.seq:*
2: /cgn2_6/prodata/2/lna/5B_COMB.seq:*
3: /cgn2_6/prodata/2/lna/5A_COMB.seq:*
4: /cgn2_6/prodata/2/lna/5B_COMB.seq:*
5: /cgn2_6/prodata/2/lna/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/lna/backfillseq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70.4	14.6	631	4	US-09-385-982-174
2	60	12.5	624	4	US-09-385-982-526
3	60	12.5	630	4	US-08-180-371-5
4	56.8	11.8	630	5	PCT-US92-05707-5
5	54.2	11.3	605	4	US-08-180-371-17
6	47	9.8	5173	1	US-09-385-982-485
7	46.4	9.6	1461	3	US-08-242-677-1
8	46.4	9.6	1461	3	US-08-722-126-4
9	46.4	9.6	1461	5	PCT-US95-04258-4
10	46.2	9.6	7218	1	US-08-232-463-14
11	45.6	9.5	1604	3	US-08-665-966-9
12	45.6	9.5	1604	3	US-09-041-780-9
13	45.4	9.4	2209	1	US-08-514-014-1
14	45.4	9.4	2209	2	US-08-833-823-1
15	45.4	9.4	2280	3	US-08-813-150-1
16	45.2	9.2	1910	4	US-08-974-691-7
17	44.4	9.2	270	2	US-08-520-678A-30
18	44.4	9.2	270	4	US-08-897-126-30
19	44.4	9.2	5555	1	US-08-484-438-3
20	44.4	9.2	19557	5	PCT-US92-06300-1
21	44.2	9.2	1046	1	US-08-361-467B-4
22	44.2	9.2	1046	1	US-08-484-332C-4
23	44.2	9.2	1117	4	US-09-247-373B-33
24	43.8	9.1	2158	1	US-07-602-608-1
25	43.8	9.1	2158	1	US-08-261-578-1
26	43.6	9.1	2674	4	US-09-817-180-1
27	43.4	9.0	960	3	US-09-248-335-57

28	43.4	9.0	1100	4	US-07-861-458C-4	Sequence 4, Appl
29	43.4	9.0	2625	4	US-09-245-041-18	Sequence 18, Appl
30	43	8.9	1577	4	US-08-821-994-59	Sequence 59, Appl
31	42.8	8.9	577	4	US-09-385-982-203	Sequence 203, Appl
32	42.6	8.9	1129	4	US-09-227-357-40	Sequence 40, Appl
33	42.6	8.9	1700	2	US-08-897-340-4	Sequence 4, Appl
34	42.6	8.9	1700	3	US-09-252-329-4	Sequence 1, Appl
35	42.4	8.8	1798	4	US-09-797-906-1	Sequence 1, Appl
36	42.4	8.8	2114	1	US-07-803-622E-6	Sequence 6, Appl
37	42.4	8.8	2628	1	US-08-143-219-1	Sequence 1, Appl
38	42.2	8.8	1493	1	US-08-340-820-24	Sequence 24, Appl
39	42.2	8.8	1493	1	US-08-593-535-24	Sequence 24, Appl
40	42.2	8.8	1921	2	US-08-557-128-11	Sequence 11, Appl
41	42.2	8.8	2550	6	5258287-23	Patent No. 5258287
42	42.2	8.8	2837	2	US-08-993-228-11	Sequence 11, Appl
43	42	8.7	880	4	US-09-247-373B-39	Sequence 39, Appl
44	42	8.7	1420	2	US-08-909-965C-3	Sequence 3, Appl
45	42	8.7	12980	3	US-08-811-566-5	Sequence 5, Appl

ALIGNMENTS

```
RESULT 1
US-09-385-982-174
Sequence 174, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCDDNA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
EARLIER FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 344
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 174
LENGTH: 631
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)..(631)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-174
```

Query Match 14.6% Score 70.4; DB 4; Length 631;
Best Local Similarity 53.7% Pred. No. 1.1e-08;
Matches 191; Conservative 0; Mismatches 161; Indels 4; Gaps 2;

QY	62	TCACACTTCCTCCCTGTATCATGTCAGATATTAAGACACGATGATTAAC	121
DB	116	tcgtaagacttcagcaagcaagatgtaagcagatgtaagcagatgctttcgcga	175
QY	122	ATTTTACAGCTGACCGACACAACTGCGATGTTCAATTTTTCGAAACGGTGG	181
DB	176	agctctggaagctgcaagtgataaacttgatgtgactctcagcaagtggtg	235
QY	182	TCCCTGCAAAAGATGTTTCTGTTTCCATGAGCTGGCTGAAACTTGTCACATCAAC	241
DB	236	gccttgaaatgataagccttcttcctcctctcgtgaaagatctcaagtgat	295
QY	242	AATGCCACATTTAGATGTTCAAGAAAGCAAGCAAGTACCTTATTCATAAATCA	301
DB	296	attccttgagatg-aggatgtagatgactgtcagatgtgtcctcaaaagtgtgaag	353

OY 302 AAGATTAATTGCTTTAGAGTGCATGATCAAGCAACTGATTTTGTGCTGG 361
Db 354 atgcatt--gccaacattccagctttttaagaagaaggacaaagtg999tgatlttcg 411
OY 362 AGCCGATGCTAAATAATGGAAGCCAGCACTCAAGATTAATTAAGTGATCTCC 417
Db 412 agccaataagaagaagcttgaaagccaccatttaataatcaatcatgttttc 467

RESULT 2

US-09-385-982-526
Sequence 526, Application US/09385982
Patent No. 6262334

GENERAL INFORMATION:

APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCNDA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
EARLIER FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 526
LENGTH: 624
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(624)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-526

Query Match 12.5%; Score 60; DB 4; Length 624;
Best Local Similarity 52.0%; Pred. No. 3.4e-06;
Matches 185; Conservative 0; Mismatches 165; Indels 6; Gaps 2;

OY 62 TCAGCACTTCCCTGCTGTAATCATGTAACGATTAATTAAGACAGATGAATTTAAAC 121
Db 65 tgcagacacacagcagcagcaagatgtgaaagcagatcgagagcagactgtttcaaga 124
OY 122 ATTTTGCAGCTGCCGACACAACTGCGAGTGTCAATTTCTGGAACGGGTGG 181
Db 125 agccttgagcgcgcagatgtaaaactgtgtagtgaactctcagcagcgtgtgtg 184
OY 182 TCCCTGCAAAAGAGATGTTCTGTTTCCATGAGCTGGCTGAACTTGCATCAAAAC 241
Db 185 gacttgcaaaatgatacagcctttcttcattccctcctgtaaaagtattccaagtgat 244
OY 242 AATAACCCACATTTGAGATGTTCAAGAAAGCCAGAGTAACCTATTTCGAAGATCAA 301
Db 245 attccttgaaatgagtgtg---gagtagctgcagagatgttctcaaggtgtgaagtc 301
OY 302 AAGATTAATTTGCTGTTATAGAGTGAATCATGAGCAACTGATTTTGAAGTTTGTGG 361
Db 302 atgctgcacaacattccgcttttttaag--aaaggacaaaagtg999tgatlttcg 358
OY 362 AGCCGATGCTAAATAATGGAAGCCAGCACTCAAGATTAATTAAGTGATCTCC 417
Db 359 agccaataagaagaagcttgaaagccaccatttaataatcaatcatgttttc 414

RESULT 3
US-08-180-371-5
Sequence 5, Application US/08180371
Patent No. 6254861
GENERAL INFORMATION:

APPLICANT: Choudhury, Chandira
TITLE OF INVENTION: Hematopoietic Growth Factor Derived Lymphocytes and Methods of Use Therefor
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESS: Marshall, O'Toole, Gerstein, Murray &
ADDRESS: Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/180,371
FILING DATE: 12-JAN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 906866
FILING DATE: 01 July 1992
ATTORNEY/AGENT INFORMATION:
NAME: Gruber, Lewis S.
REGISTRATION NUMBER: 30,060
REFERENCE/DOCKET NUMBER: 27620/31668
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 630 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..630
US-08-180-371-5

Query Match 12.5%; Score 60; DB 4; Length 630;
Best Local Similarity 52.0%; Pred. No. 3.5e-06;
Matches 185; Conservative 0; Mismatches 165; Indels 6; Gaps 2;

OY 62 TCAGCACTTCCCTGCTGTAATCATGTAACGATTAATTAAGACAGATGAATTTAAAC 121
Db 204 TCGTACAGCTCCAGCAGCAAGATGTAAGCAAGATGAGAGCAAGACTGTTTCAAGA 263
OY 122 ATTTTGCAGCTGCCGACACAACTGCGAGTGTCAATTTCTGGAACGGGTGG 181
Db 264 AGCTTGACCTCCAGCTGATTAACCTGTAAGTGTGACTTCTCAGCACTGCTGG 323
OY 182 TCCCTGCAAAAGAGATGTTCTGTTTCCATGAGCTGGCTGAACTTGCATCAAAAC 241
Db 324 GCTTGGCAAAATGATCAAGCTTTCTTCAATCCCTCTCTGTAAGATTTCCACAGTAT 383
OY 242 AATAACCCACATTTGAGATGTTCAAGAAAGCCAGAGTAACCTATTTCGAAGATCAA 301
Db 384 ATTCCTGAAATAGATG---GATGACTGTCAGAGATGTTCTTCAAGATGTAAGTCAA 440
OY 302 AAGATTAATTTGCTGTTATAGAGTGAATCATGAGCAACTGATTTTGAAGTTTGTGG 361
Db 441 ATGCATGCCAACATTCAGTTTTTAAG--AAAGGACAAAAGGTGGGTGATTTTCTGG 497
OY 362 AGCCGATGCTAAATAATGGAAGCCAGCACTCAAGATTAATTAAGTGATCTCC 417
Db 498 AGCCATTAAGAAAGCTTGAAGCCACCATTAATTAATTAATTAATTAATTAATTAATTA 553

RESULT 4
PCT-US92-05707-5
Sequence 5, Application PC/TUS9205707
GENERAL INFORMATION:
APPLICANT: Choudhury, Chandra
TITLE OF INVENTION: Hematopoietic Growth Factor Derived from
TITLE OF INVENTION: T Lymphocytes and Methods of Use Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Bicknell
STREET: Two First National Plaza, 20 South Clark
STREET: Street
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/05707
FILING DATE: 19920707
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Gruber, Lewis S.
REGISTRATION NUMBER: 30,060
REFERENCE//DOCKET NUMBER: 27620/30933
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 630 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..630
PCT-US92-05707-5

Query Match 12.5%; Score 60; DB 5; Length 630;
Best Local Similarity 52.0%; Pred. No. 3.5e-06;
Matches 185; Conservative 0; Mismatches 165; Indels 6; Gaps 2;

QY 62 TCAGCACTTCCCTGGTGTATCATGTACAGATTATTAAGACGAGTGAATTTAAAC 121
DB 204 TCGCAACACTCCACACCAAGATGTGAAGCAGATGAGAGCAAGCTGTTTCAGGA 263
QY 122 ATTTTGAACGTCGCGACAACTCGAGTGTCAATTTCTTCGAAACGGTGG 181
DB 264 AGCCTTGAGCGCTGAGGTATTAACCTGTAGTAGTTGACTTCAGCGCAGTGTGG 323
QY 182 TCCCTGCAAAAGATGTTCTGTTTCATGAGCTGGCTGAACCTGTGCATCAAAAC 241
DB 324 GCCTTGCAAAATGATCAAGCTTCTTCATTCCTCTCTCAAAAGATTCACAGTGAT 383
QY 242 AATACCAATTTGATGTTTCAAGAAAGCAGAGTAACTTATTCAGAAATCAA 301
DB 384 ATTCTTGAGTAGATGTG---GATGACTGTCAGAGATGTGCTTCAGAGTGTGAATCAA 440
QY 302 AAGATAATTTGCTGTATAGAGTTCATGAGCAACTGATTTTGTGTTTGG 361
DB 441 ATGATGCCAATTCATCCAGTTTTTAAG---AAGGACAAAGAGTGGTGAATTTTTCGG 497
QY 362 AGCGATGTATAAAATTTGAAGCAAGACTCAAGAAATTAATTAAGCTGATCTCC 417

Db 498 AGCAATAGGAAAGTTGAGCCACCATTAATGATTAAGTATGATCATGCTTTC 553

RESULT 5
US-08-180-371-17/c
Sequence 17, Application US/08180371
Patent No. 6254861
GENERAL INFORMATION:
APPLICANT: Choudhury, Chandra
TITLE OF INVENTION: Hematopoietic Growth Factor Derived
TITLE OF INVENTION: from T Lymphocytes and Methods of Use Therefor
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/180,371
FILING DATE: 12-JAN-1994
CLASSIFICATION: 424
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: US 906866
FILING DATE: 01 July 1992
ATTORNEY/AGENT INFORMATION:
NAME: Gruber, Lewis S.
REGISTRATION NUMBER: 30,060
REFERENCE//DOCKET NUMBER: 27620/31668
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 630 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-180-371-17

Query Match 11.8%; Score 56.8; DB 4; Length 630;
Best Local Similarity 51.4%; Pred. No. 2.1e-05;
Matches 183; Conservative 0; Mismatches 167; Indels 6; Gaps 2;

QY 62 TCAGCACTTCCCTGGTGTATCATGTACAGATTATTAAGACGAGTGAATTTAAAC 121
DB 427 TCGTCAAGACTCCACAGCAAGATGTGAAGCAATGAGAGCAAGCTGTTTCAGGA 368
QY 122 ATTTTGAACGTCGCGACAACTCGAGTGTCAATTTCTTCGAAACGGTGG 181
DB 367 AGCCTTGAGCGCAGAGGTATTAACCTGTAGTAGTTGACTTCAGCGCAGTGTGG 308
QY 182 TCCCTGCAAAAGATGTTCTGTTTCATGAGCTGGCTGAACCTGTGCATCAAAAC 241
DB 307 GCCTTGCAAAATGATCAAGCTTCTTCATTCCTCTCTCAAAAGATTCACAGTGAT 248
QY 242 AATACCAATTTGATGTTTCAAGAAAGCAGAGTAACTTATTCAGAAATCAA 301
DB 247 ATTCCTGAGTAGATGTG---GATGACTGTCAGAGATGTCTTCAGAGTGTGAATCAA 191
QY 302 AAGATAATTTGCTGTATAGAGTTCATGAGCAACTGATTTTGTGTTTGG 361


```

TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1461 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 54...617
PCT-US95-04258-4

Query Match          9.6%; Score 46.4; DB 5; Length 1461;
Best Local Similarity 70.5%; Pred. No. 0.0081;
Matches 62; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 394 AAGATTATATGTAAAGCTGATCTCCCAAGGCCAAATACACTTGTGACATTTGAAAGGCCAAA 453
      || || || || || || || || || || || || || || || || || || || || || ||
Db 1347 AAAAATTAAAGAGCCCAATTAATAATAAAATACCTTTCTGCTAAAAAATAAAAAAAAAA 1406

QY 454 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 481
      || || || || || || || || || || || || || || || || || || || || || ||
Db 1407 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434

```

RESULT 10
 US-08-232-463-14/c
 : Sequence 14, Application US/08232463
 Patent No. 5670367
 GENERAL INFORMATION:
 APPLICANT: DORNER, F.
 APPLICANT: SCHIEFFLINGER, F.
 APPLICANT: FALKNER, F. G.
 TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
 NUMBER OF SEQUENCES: 52
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 1800 Diagonal Road, Suite 500
 CITY: Alexandria
 STATE: VA
 COUNTRY: USA
 ZIP: 22313-0299
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/232,463
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/935,313
 FILING DATE:
 APPLICATION NUMBER: EP 91 114 300.6
 FILING DATE: 26-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 30472/114 IMMU
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)836-9300
 TELEFAX: (703)683-4109
 TELEX: 899149
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7218 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single

APPLICANT: Liu, Yong-Jun
APPLICANT: Dowling, Lynette M.
APPLICANT: Huffine, Constance M.
APPLICANT: Gorman, Daniel M.
TITLE OF INVENTION: MAMMALIAN PROTEINASES; OXIDOREDUCTASES;
RELATED REAGENTS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: DNAX Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/813,150
FILING DATE: 07-MAR-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: SF0693
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-852-9196
TELEFAX: 650-496-1200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2280 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 61..1470
US-08-813-150-1

Query Match 9.4%; Score 45.4; DB 3; Length 2280;
Best Local Similarity 67.4%; Pred. No. 0.016;
Matches 64; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 387 AAGACTCAAGATTAATGTAAGCTGATCTCCAGGCAAAATATCACTGTGACATTGAA 446
DB 2134 AAGGCTTAATGTTATGTAACCATTAATGTCATTAATAACACACACTTTTGAAGCAAAA 2193
QY 447 AGGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 481
DB 2194 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2228

Search completed: June 8, 2002, 03:21:07
Job time: 3947 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2002, 02:09:55 ; Search time 1800.24 Seconds
(without alignments)
3606.206 Million cell updates/sec

Title: US-09-746-783-18

Perfect score: 481
Sequence: 1 GGATCCTGTAATAATGGA.....AAAAAAAAAAAAAAAAAAAA 481

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 segs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estf2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_hiv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	149.6	31.1	399	9	A1188241
2	144.6	30.1	619	11	AK015240
3	144.2	30.0	491	11	AK006683
4	141.4	29.4	526	9	AV209695
5	126.4	24.5	344	9	A1764117
6	118	22.7	245	9	AV045293
7	109.4	22.7	266	9	AV269460
8	96	20.0	666	12	AG084227
9	85	17.7	286	9	AV269811
10	83.8	17.4	384	10	BF149520
11	83.6	17.4	428	9	A1885805
12	75.2	15.6	502	12	AQ351522
13	73.4	15.3	907	10	BF028142
14	72.8	15.1	273	9	AV267588
15	68.6	14.3	846	10	BF028467
16	67.6	14.1	460	10	Z71861
17	67.4	14.0	849	10	BF210792

18	67.2	14.0	424	10	W87705	W87705 zh55c10.r1
19	66.4	13.8	468	10	W95350	W95350 ze06b03.r1
20	65.4	13.6	459	10	BF109086	BF109086 7150407.x
21	65.2	13.6	565	9	A1815810	A1815810 au42f03.x
22	65.2	13.6	569	10	BG567720	BG567720 602586510
23	64.8	13.5	763	10	BE787457	BE787457 601479729
24	64.6	13.4	546	10	BF970488	BF970488 602272657
25	64	13.3	455	10	W69590	W69590 zd47c02.r1
26	63.8	13.3	426	10	BG937432	BG937432 1AB01C11
27	63.8	13.3	427	10	BM429932	BM429932 1DU03C8.
28	63.8	13.3	424	10	BM433322	BM433322 10m12E09
29	63.8	13.3	516	9	AV663283	AV663283 AV63496
30	63.8	13.3	561	9	AV593496	AV593496 AV593496
31	63.8	13.3	603	10	BG027655	BG027655 602294668
32	63.4	13.2	560	10	BF246611	BF246611 601854681
33	63.2	13.1	393	9	AJ275295	AJ275295 AJ275295
34	63.2	13.1	455	9	AW429592	AW429592 67920 MAR
35	63.2	13.1	472	10	B1359967	B1359967 384756 MA
36	63.2	13.1	522	9	A1620253	A1620253 lu93b03.x
37	63.2	13.1	544	10	BF218303	BF218303 601881647
38	63.2	13.1	575	10	B1346139	B1346139 375177 MA
39	63	13.1	501	9	AA810660	AA810660 0D58E04.S
40	63	13.1	546	10	BE967335	BE967335 601649331
41	62.8	13.1	397	10	BE440038	BE440038 HTM1-809F
42	62.8	13.1	458	9	AV593495	AV593495 AV593495
43	62.6	13.0	485	10	N77546	N77546 YZ89B11.r1
44	62.4	13.0	461	10	W76348	W76348 zd66d02.r1
45	62.4	13.0	503	9	AA045052	AA045052 ZK76912.8

ALIGNMENTS

RESULT 1
A1188241/c
LOCUS
DEFINITION
gq67d05.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1734537
3' similar to SW:THIO_HUMAN P10599 THIOREDOXIN ; mRNA sequence.

ACCESSION
A1188241
VERSION
A1188241.1 GI:3739450
KEYWORDS
EST.

SOURCE
ORGANISM
human.

REFERENCE
AUTHORS
TITLE
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
1 (bases 1 to 399)
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL
CONTACT: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.C.E. Consortium/LNL at:
www.bio.linn.gov/bdrp/image/image.html
Insert length: 523 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 347.

FEATURES

source
1..399
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1734537"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT773D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech laboratories

BASE COUNT 120 a 83 c 75 g 121 t
 ORIGIN

Query Match 31.1%; Score 149.6; DB 9; Length 399;
 Best Local Similarity 75.5%; Pred. No. 3.7e-10;
 Matches 228; Conservative 0; Mismatches 14; Indels 60; Gaps 1;

QY 126 TTGACAGCTGGCGACACAACTCCAGTGGTTCATTTTCTCGAAACGGTGTGCTCC 185
 |||||
 Db 399 TTGACAGCTGGCGACACAACTCCAGTGGTTCATTTTCTCGAAACGGTGTGCTCC 340
 |||||

QY 186 TGCAGAAAGGATGTTCTGTTTTCAT-----212
 |||||
 Db 339 TGCAGAAAGGATGTTCTGTTTTCAT-----212
 |||||

QY 213 -----GAGCTGGCTGAACCTGTCACATCAAAACATA 245
 |||||
 Db 279 GCTAATGTGATGTGACAAATTCCTCCGAGCTGCTGAACCTGTCACATCAAAACATA 220
 |||||

QY 246 CCCAATTTTCAGATGTTTCAGAAAGCCAGAAAGTACCTATTTCTCAAGATCAAAAGA 305
 |||||
 Db 219 CCCAATTTTCAGATGTTTCAGAAAGCCAGAAAGTACCTATTTCTCAAGATCAAAAGA 160
 |||||

QY 306 ATAAATTTCTGTATAGAGTGATGATTCAGCAACCTGATTTTGAATTTTGTGGAGCC 365
 |||||
 Db 159 ATAAATTTCTGTATAGAGTGATGATTCAGCAACCTGATTTTGAATTTTGTGGAGCC 100
 |||||

QY 366 GA 367
 ||
 Db 99 GA 98

RESULT 2
 AK015240 619 bp mRNA linear HTC 19-JAN-2002
 LOCUS
 DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930429J24;homolog to THIOREDOXIN, full insert sequence.
 ACCESSION AK015240
 VERSION AK015240.1 GI:12853498
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (Strain:C57BL/6J) adult male testis cDNA to mRNA, clone_4930429J24.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (sites)
 AUTHORS Carninci,P. and Hayashizaki,Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE 2 (sites)
 AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3 (sites)
 AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsuami,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fuchida,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matshiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)

4 (sites)
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5 (bases 1 to 619)
 Aizawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schirml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
 Direct Submission
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehico-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gs.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
 Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGATTCGACGATTAATTAATTAATCCCCCCCCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot - 10.0 and subtraction to Rot - 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGACGATTAATTAATTAATCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from lambdaBda FLC I. Cloning sites, 5' end: SalI, 3' end: BamHI. Host: DH10B.
 Location/Qualifiers
 1. 619
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="MGD:GI:1896676"
 /db_xref="taxon:10090"
 /clone="4930429J24"
 /sex="male"
 /tissue-type="testis"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev-stage="adult"
 97. 480
 /note="data source:SPTR, source key:P08628, evidence:ISS homolog to THIOREDOXIN
 putative"
 /codon_start=1
 /protein_id="BAB29760.1"
 /db_xref="GI:12853498"
 /translation="MVRIRKMSLEKLFSDAGNKLVVFEFSKACGPGCTIAPVDA


```

/codon_start=1
/protein_id="BAB24702.1"
/db_xref="GI:12839899"
/translation="MVKRIKNNSEKLEKELSDAGNKLIVVEFSAKMGCPCKTIAPFOA
MSLKVONTFAOVDSKSKLEAHCODITMLPTFPMFKYTKQVTFPSRIKRVLCCLRSR
PKSMIFECHEADKKOLEKTIQELM"
BASE COUNT      157 a      102 c      115 g      117 t
ORIGIN

```

```

Query Match      30.0%; Score 144.2; DB 11; Length 491;
Best Local Similarity 64.2%; Pred. No. 1.5e-09;
Matches 300; Conservative 0; Mismatches 103; Indels 64; Gaps 3;

```

```

Oy 17 AGGAGACAGCTACAGTATCC---AAGTAACCAACAGGGGATTTTCATCAGCACTCC 73
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5  ATGGACAGCTACAGTATCCCTCCACCTGAACCAACAGGAGGATGACAGCAAGTCCC 64
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 74 TGGTG-TAATCATGGTACAGATTATTAAAGACAGATGAATTAAACATTTTGACAG 132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 AGATTTTATCATGTGTGAAGAGATTAAAGACATGATGAAGTAAGTAAGTATGTCACG 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 133 CTGCGGACACAAACTCGAGTGTTCATTTTCGAAAGCGTGTGCTCCGCAAAA 192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 ATGCTGGAACAAAGCTTGTGGTGAAGTTTTCAGCAAGTGTGTGCCCCCTCAAAA 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 193 GGATCTTCTGTTTTCATGAGCTG----- 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 185 CAATTGCTCTGTTTTCAGAGCATGCTTTAAATACCAAAATGTCAGTGTGCTCAGG 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 219 -----GCTGAACCTGTACATCAAAACAAATACCAAT 252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 245 TGGATGTGACTCATCTAAGCAATTAGCTGAACATTTGACATCAAGTCAACCAAT 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 253 TTCAGATGTCAGAAAGCCAGAGTAACCTATTCAGAAATCAAAAGATATATTT 312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 TCCAGATGTCAGAGTACCCAAAGTCACTCCATTCAGATCAAAAGATAGTCT 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 313 GCTGTATAGAAAGTGAATTCATGAGCAACCTGATTTTGGATTTTGGAGCGGATG 372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 GCTGCTCAGAGATGAGCCAGAGCAAGATGATTTTGAATGTCATGAGCTGATGCTA 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 373 AAAAATGGAAGCCAGAGTCAAGAAATTAATGTAAGTGAATCTGATCTCCAA 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 AAACAAGTGAAGAAAGATTCAGAGCTAATGTAAACGCCACCCAA 471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 4
AV209695 526 bp mRNA linear EST 24-OCT-2001
LOCUS AV209695 RIKEN full-length enriched, adult male testis Mus musculus
DEFINITION CDNA clone I700120N16 3', mRNA sequence.
ACCESSION AV209695.2 GI:16384508
VERSION AV209695
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.

```

```

REFERENCE
AUTHORS Eukaryota, Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
1 (bases 1 to 526)
Atakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanazaki, T., Hara, A.,
Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Kono, H., Kouda
M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki
D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,
Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Atakawa, T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Oct 29, 1999 this sequence version replaced gi:6150401.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute

```

The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9220
 Fax: 81-45-503-9216
 Email: genome-res@gscc.riken.go.jp
 URL: <http://genome-gsc.riken.go.jp/>
 Carninci, P., Shibata, Y., Hayasaka, M., Sugahara, Y., Shibata, K., Itoh M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Itawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillarary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)
 Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara Y., and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)
 Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamakawa, I., Aizawa K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.
 Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. *Genome.* 12, 673-677 (2001)
 Please visit our web site (<http://genome-gsc.riken.go.jp/>) for further details.
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Location/Qualifiers
 1..526
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="I700120N16"
 /clone_lib="RIKEN full-length enriched, adult male testis"
 /sex="male"
 /tissue-type="testis"
 /dev_stage="adult"
 /lab_host="SODR"
 /note="Site.1: XhoI; Site.2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGGAGAGAGAGATCCAGAGCTCTTTTGTGTTTATTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGGAGAGAGAGCGCGCCCAATTAATTCAGATTAATTAATTAATTCGCC 3']. cDNA was cloned into the XhoI and BamHI sites."

```

BASE COUNT      162 a      122 c      117 g      119 t
ORIGIN
Query Match      29.4%; Score 141.4; DB 9; Length 526;
Best Local Similarity 63.5%; Pred. No. 3.3e-09;
Matches 283; Conservative 0; Mismatches 102; Indels 61; Gaps 2;
Oy 35 TCCAACTAACAACAGGAGATTTTCATCAGCACTCCCTGGTG-TAATCATGTGACAGA 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 TCCCTGTAACAACAGAGAGGAGGTGACAGCAAGTCCAGATGATGATGATGTAAGA 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 94 TTATTTAAAGACAGAAATGAATTTAAACATTTTTCAGACAGCTGCGGACACAACTCGAC 153

```

```

121 GGAATTAAGACATGAGTGAAGTGAAGAAATGTTACGATGCTGGAACAAAGCTTGCG 180
154 TGGTCAATTTCTTCGAAAGCGTGTGCTCCGCAAAAGGATGTTCTGTTTCATG 213
181 TGGTAGAGTTTTCAGCAAGTGGTGTGCCCCCTGCAAAACAAATGCTCTCTTTCCAGG 240
214 AGCTG----- 218
241 CAATGCTTTAAATACCAAAATGTCAAGTTTCTCAGTGATGATGGACTCATTAAG 300
219 -----GCTGAACCTGTCACATCAAAACATACCCACATTCAGATGTTCAAGAAAGCC 273
301 AATTAGCTGAACATTTGACATCNCATGCAATGCAATTCACATTCGATTCAGATCACACC 360
274 AGAAGTAACCCCTTCTTCAGAAATCAAAAGAAATATTTGCTTTATAGAGTGAATTC 333
361 AAAAGTCACCTCCATTCCTCAGACTCAAAAGAGTGTGCTGCTCCTCAGAGTGGACCGA 420
334 TGAACAACCTGATTTTGTGAGTTTGTGAGCCGATGCTAAAAAATTTGGAAGCAAGACTC 393
421 AGAGCAAGATGATTTTGAATGTCATGAGCTGATGCTAAACAACTGGAAGAAAGATTC 480
394 AAGATTAATGTAAGCTGATCTCGAA 419
481 AAGAGTAATGTAAGCAAGCCCA 506

```

```

RESULT 5
A1764117/c 344 bp mRNA linear EST 25-JUN-1999
LOCUS A1764117
DEFINITION UI-R-Y0-acr-a-07-0-UI.s1 UI-R-Y0 Rattus norvegicus cDNA clone
VERSION A1764117 GI:5210052
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 344)
REFERENCE 1
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery.
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized Eye library cDNA library preparation: M.B. Soares Lab
Clone distribution: clones will be available through Research
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=yes.

```

```

FEATURES
Source Location/Qualifiers
1. 344
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-Y0-acr-a-07-0-UI"
/clone_lib="UI-R-Y0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"

```

```

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-Y0
library is a subtracted library derived from an
individually-tagged normalized whole-eye (minus the lens)
library. The driver for the subtraction consisted of a
pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,
UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
3-5 nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-Y0) was constructed as follows: PCR
amplified cDNA inserts from previous library clones from
which 3' ESTs had been derived were used as a driver in a
hybridization with the normalized whole-eye library in
the form of single-stranded circles. The remaining
single-stranded circles (subtracted library) was purified
by hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-Y0
library. This procedure has been previously described
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
1996)
TAG_LIB=UI-R-Y0
TAG_TISSUE=Eye
TAG_SEQ=CATTG"
BASE COUNT 90 a 70 c 69 g 115 t
ORIGIN

```

```

Query Match 26.3%; Score 126.4; DB 9; Length 344;
Best Local Similarity 69.4%; Pred. No. 3,16-07;
Matches 172; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

```

```

QY 210 CATGAGCTGCGTGAACCTTGACATCAAAACATACCCACATTCATGTTCAAGAA 269
DB 248 CAGGATGAGCTGAACATGTTGATGATCCAGTAGGCGCATTCACATGCTCAAGC 189
QY 270 AGCCGAGGATGATCCCTATTCAGATCAAAAGAAATATTTGCTTATGAGTGA 329
DB 188 TCCGGAAGAGTCACTCCATTCATCAAGACTCAAAAGAAATACGTGCTGTTCAAGTGA 129
QY 330 TTCAAGACCAACCTGATTTTGTGAGTTTGTGAGCCGATGCTAAAAAATTTGGAAGCC 389
DB 128 CCAGGAGCAGAAAGATTTTGAATTTCAAGCCGCTGATTTGAAAAATCGAAGAGA 69
QY 390 ACTCAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 449
DB 68 ATTCAGAGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 9
QY 450 CAAAAAAA 457
DB 8 AAAAAAAA 1

```

```

RESULT 6
AV045293 245 bp mRNA linear EST 23-NOV-1999
LOCUS AV045293
DEFINITION AV045293 Mus musculus adult C57BL/6J testis Mus musculus cDNA clone
VERSION AV045293.2 GI:4864958
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE 1
AUTHORS Carninci, P., Shibata, K., Ozawa, Y., Kono, H., Itoh, M., Aizawa, K.,
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara,
A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Iwasa, M., Kawai, J.,
Kikuchi, N., Kojima, Y., Matsuyama, T., Natsushima, H., Oda, H., Owa, C.,
Sato, K., Shibata, Y., Shigemoto, Y., Shitaki, T., Sogabe, Y., Suganara,
Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N.,

```


ACCESSION	musculus cdna clone 4930545G19 3', mRNA sequence.
VERSION	AVZ69811
KEYWORDS	AVZ69811.1 GI:6257848
SOURCE	EST.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 296) Konno,H., Aizawa,K., Akahira,S., Akiyama,U., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai .C., Kawai,T., Kikuchi,N., Kojima,Y., Koya,S., Kusabebe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata .Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugihara,Y., Suzuki,H., Suzuki,K., Takahashi,F., Tateo,M., Tomiyaga,N., Tsunoda,Y., Watanabe,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE	RIKEN Mouse ESTs (Konno,H., et al. 1999)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Shuhro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@sc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Sasaki,N., Izawa,M., Wataniki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki .Y. Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7) 3455-3460 (1998) Itoh,M., Kitsunai,T., Akiyama,U., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki .Y. and Hayashizaki,Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 13-44 (1999) Please visit our web site (http://genome.ritc.riken.go.jp) for further details.
FEATURES	Location/Qualifiers
SOURCE	1..296 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="4930545G19" /clone_lib="RIKEN full-length enriched, adult male testis (DH10B)" /sex="male" /issue_type="testis" /dev_stage="adult" /lab_host="DH10B" /note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGAGAAGAGATCCAAAGAGCTCTTTTCTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAAGATTCGAGTTAATTAAATTATATCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

BASE COUNT	90 a	67 c	63 g	76 t
Query Match:	17.7%;	Score 85;	DB 9;	Length 296;
Best Local Similarity	67.5%;	Pred. No. 0.049;		
Matches 137;	Conservative 0;	Mismatches 60;	Indels 6;	Gaps 1;
OY	257	GATGTTCAAGAAACCCAGAGTAACTACCTATTCTCCAGATCAAAAGCATATTGCTG	316	
Db	1	GATCTTCAACTACACCCAAAAGTTTACCTCAATTTTCAAGACACCAAGGAAGTGTGATG	60	
OY	317	TTATATGAGTGATTCATGAGCAACCTCATTTTGATGTTTGAGAGCGATGCTAAAAA	376	
Db	61	CCTCAGAAAGTGCACCCGAAAGACCAATATGTTTTTAAAGTCATGAGGCTGATGCTCAACA	120	
OY	377	ATTGAGACCAAGACTCAAGAAATATGATGATCTCCAAAGCCAAATACACTGTG	436	
Db	121	ACTGAAAAAAGAGATTCAAGACTAATGTAAAGGCCAACCC-----AATATACTGTG	174	
OY	437	ACATTGAAAGGCGCAAAAAA	459	
Db	175	ACATTTCCTAAAGCAGAGCTAA	197	
RESULT 10				
LOCUS	BF149520			
DEFINITION	BF149520	384 bp	mRNA	linear
ACCESSION	BF149520			
VERSION	BF149520.1	GI:11030915		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 384)			
AUTHORS	Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.			
TITLE	The WashU-NCI Mouse EST Project 1999			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Marra M/WashU-NCI Mouse EST Project 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:1425258 Seq primer: Primer name ambiguous High quality sequence stop: 368. Location/Qualifiers 1. 384 /organism="Mus musculus" /strain="CD-1" /db_xref="taxon:10090" /clone="IMAGE:3664490" /clone_id="McCarrey Eddy round spermatid" /sex="male" /tissue_type="round spermatids, pooled from multiple mice" /dev_stage="60 day" /lab_host="DHI08 (phage-resistant)" /note="Organ: testis; Vector: pBluescript SK+ (Stratagene [5']-(GA)10-ACTAGTCGAGTGTGTTT-3') and directionally cloned using 5' linkers 5'-ATTGGCAGCAG-3' and 5'-CTGTCGCCG-3'. Size selection of >400bp material gives			

		average insert size ranging from 1-2 kb. Library was mass excised (from lambda-unizap-XR) and resulting single-stranded phagemids were prepped and transformed into DH10B. Library contains 98.5% recombinants. References: J. Androl. 20:635-639 and Gene 25:263-269. Library constructed and donated by J. McCarrey, Ph.D. (Southwest Foundation for Biomedical Research, Dept. of Genetics); excision done by E.M. Eddy, Ph.D. (National Institutes of Health, National Institute of Environmental Health Sciences). Original lambda-based library is available through ATCC, catalog #63423."		
BASE COUNT	117 a	86 c	90 g	91 t
ORIGIN				
Query Match	17.4%;	Score 83.8;	DB 10;	Length 384;
Best Local Similarity	76.3%;	Pred. No. 0.059;		
Matches 103;	Conservative	0;	Mismatches 32;	Indels 0; Gaps 0;
OY	213	GAGCTGGCTGAACCTGTGCATCAAAACAATACCACATTTCAGATGTTCAAGAANNC	272	
Db	250	GAATTAGCTGAACATTTTGATCATCATACATGCTACCCACATTCAGATGTTCAACTACAC	309	
OY	273	CAGAAGGTAAACCTATTCTCAGAAATCAAAGATTAATTGCTGTTATAGAGTGGATTC	332	
Db	310	CAAAAGGTCACTCCATTCTCAAGACTCAAAGAGTACTGTCCTCTCAGAAAGTAGCACCG	369	
OY	333	ATGAGCAACCTGATT	347	
Db	370	AAGACAGATGATT	384	
RESULT 11				
AI885805/c		428 bp	mRNA	linear
LOCUS	wI61g04.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2429430 3'			
DEFINITION	s similar to gb:X77564 THIOREDOXIN (HUMAN).; mRNA sequence.			
ACCESSION	AI885805			
VERSION	AI885805.1			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 428)			
JOURNAL	NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.			
COMMENT	National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTCAP), Tumor Gene Index Unpublished (1998)			
	Contact: Robert Strausberg, Ph.D. Email: cgapsb@emall.nih.gov Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LBL at: www.bio.lnl.gov/dbfp/image/image.html Seq primer: -400P from Gibco High quality sequence stop: 40A. Location/Qualifiers			
FEATURES	location/Qualifiers			
source	l.428			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone_lib="IMAGE:2429430"			
	/clone_lib="NCI_CGAP_Brn25"			
	/tissue_type="anaplastic oligodendroglioma"			
	/lab_host="DH10B"			
	/note="Organ: Brain; Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st			

> 0 <
0110 Intelligenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file Seq19.res made by tport on Mon 24 Jun 102 8:34:12-PDT.

Query sequence being compared: US-09-746-783-19 (1-107)
Number of sequences searched: 1
Number of scores above cutoff: 1

Results of the initial comparison of US-09-746-783-19 (1-107) with:
File: p10599.pep

```

100-  -
N    -
U    -
M    -
B    -
E    -
R    -
O    -
F    -
S    -
E    -
D    -
U    -
N    -
C    -
E    -
S    -
SCORE 0 4 8 13 17 21 25 30 34 38
STDEV

```

PARAMETERS

Similarity matrix PAM-150 K-tuple 1
Threshold level of sim. 16%
Mismatch penalty 1 Joining penalty 20
Gap penalty 5.00 Window size 107
Gap size penalty 0.05
Cutoff score 1
Randomization group 0

SEARCH STATISTICS

Scores: Mean 38 Median 0 Standard Deviation 0.00
Times: CPU 00:00:00.00 Total Elapsed 00:00:00.00
Number of residues: 105
Number of sequences searched: 1
Number of scores above cutoff: 1

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Score	Opt. Frame
1. thiohuman	TOIG of: thio_human	check: 3	105	38 61 0.00 0

1. US-09-746-783-19 (1-107)
thiohuman TOIG of: thio_human check: 3392 from: 1 to: 105
Initial score = 38 Optimized score = 61 Significance = 0.00
Residue Identity = 31% Matches = 34 Mismatches = 65
Gaps = 2 Conservative Substitutions = 6

```

X      10      20      30      40      50      60      70
MVOIITKTNEFKTFLTAAGHKLAVVQSSKRCGPCKRMFPVPHLEAETCHIKTIPTFQMRKKSOKVTLERSRI
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
MVKQIESKTAFQDALDAAGDKLVVWDFSATWCGPCMKIKPFPHSLSEKY--SNVIFLEVDYDDCODVASCE
X      10      20      30      40      50      60      70
KRIICYRSGFMSNLIFEECGADAKKILEAKTOELM
:      :      :      :      :      :      :      :      :      :      :      :      :
VKCMPTFOFPFKKGQKVGEGESGANKKEKLEATINELY
80      90      100      X
80      90      100      X

```


[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2002, 04:29:56 ; Search time 36.79 Seconds

(without alignments)
112.612 Million cell updates/sec

Title: US-09-746-783-19

Perfect score: 562
Sequence: 1 MWQIKDTRNEFKFLTAAGH.....IFEFGADAKKLEAKTQELM 107

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	185	32.9	104	THIO_HUMAN	P10599 Homo sapien
2	185	32.9	104	THIO_SHEEP	P50413 ovine aries
3	184	32.7	104	THIO_RABIT	P08628 oryctolagus
4	182	32.4	104	THIO_BOVIN	O97680 bos taurus
5	180	32.0	104	THIO_MOUSE	P10638 mus musculu
6	179	31.9	104	THIO_RAT	P11237 rattus norv
7	177	31.5	104	THIO_MACMU	P29451 macaca mola
8	160	28.5	104	THIO_CHICK	P08629 gallus gall
9	124	22.1	106	THIO_COPCM	O94002 coprinus co
10	118	21.0	289	TXNL_HUMAN	O43396 homo sapien
11	111	19.8	102	THIO_CHPN	O92798 chlamydia p
12	110	19.6	102	TRX2_SCHPO	O14463 schizosacch
13	110	19.6	102	TRX1_SCHPO	P58465 schizosacch
14	106.5	19.2	102	THIF_PEA	P29450 pisum sativ
15	103.5	18.4	190	THIF_SPIOI	P09856 spinacia ol
16	101.5	18.1	107	THIO_DROME	P47938 drosophila
17	101.5	18.1	185	THF2_ARATH	O6xfn8 arabidopsis
18	100.5	17.9	133	THH2_ARATH	O38879 arabidopsis
19	100.5	17.9	182	THIF_BRANA	O48897 brassica na
20	99	17.6	105	THI1_DICDI	P29445 dictyosteli
21	98.5	17.5	178	THF1_ARATH	O9xfn8 arabidopsis
22	98	17.4	102	THIO_CVAME	O32022 cyanidiosi
23	98	17.4	104	THI3_DICDI	P29447 dictyosteli
24	97.5	17.3	103	TRX1_YEAST	P22803 saccharomyc
25	96.5	17.2	191	THIF_MESCR	O61333 mesembryan
26	95	16.9	102	THIO_CHLMU	O9pjks chlamydia m
27	94	16.7	112	THIO_CHLRE	O44544 chlamydia t
28	92.5	16.5	112	THIO_CHLRE	P80028 chlamydomon
29	92	16.4	127	THIO_CHRPS	P52227 chlamydia p
30	92	16.4	127	THIO_NEUCR	P42115 neurospora
31	90.5	16.1	122	THI2_ORYSA	O42443 oryza sativ
32	89	15.8	88	THI2_DICDI	P29446 dictyosteli
33	89	15.8	126	THI1_WHEAT	O64394 triticum ae

34	88.5	15.7	107	1	THIO_CVACA	P37395 cyanidium c
35	86	15.3	118	1	THH5_ARATH	O39241 arabidopsis
36	86	15.3	175	1	THIM_WHEAT	O92p21 triticum ae
37	85.5	15.2	106	1	THIO_PENCH	P34723 penicillium
38	85.5	15.2	114	1	THH1_ARATH	P29448 arabidopsis
39	85.5	15.2	172	1	THIM_PEA	P48384 pisum sativ
40	84.5	15.0	156	1	DYL5_CHLRE	O39592 chlamydomon
41	84.5	15.0	519	1	PDAS_HUMAN	O14554 homo sapien
42	84	14.9	102	1	TRX2_YEAST	P22217 saccharomyc
43	84	14.9	109	1	THIO_YEAST	P29429 emericocella
44	83.5	14.9	115	1	THIO_MYCTU	P52228 mycobacteri
45	83	14.8	115	1	THIO_CABEL	O09433 caenorhabdi

ALIGNMENTS

RESULT	1	THIO_HUMAN	STANDARD:	PRT:	104 AA.
AC	P10599:	THIO_HUMAN			
DT	01-JUL-1989 (Rel. 11, Created)				
DT	01-DEC-1992 (Rel. 24, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Thioredoxin (ATL-derived factor) (ADF) (Surface associated sulphhydryl protein) (SASP).				
GN	TXN OR TRDX OR TRX.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
ON	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=91340156; PubMed=1874447;				
RT	Tonissen K.F., Wells J.R.E.;				
RL	"Isolation and characterization of human thioredoxin-encoding genes.";				
	Gene 102:221-228(1991).				
ON	[2]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=89008454; PubMed=3170595;				
RT	Woliman E.E., D'Aurion L., Rimsky L., Shaw A., Jacquot J.-P.,				
RL	Wingfield P., Graber P., Dessarps F.;				
RT	"Cloning and expression of a cDNA for human thioredoxin.";				
RL	J. Biol. Chem. 263:15506-15512(1988).				
ON	[3]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=89251607; PubMed=2785919;				
RT	Tagaya Y., Maeda K.-I., Mitsui A., Kudo N., Matsui H., Hamuro J.,				
RL	Brown N., Arai K.-I., Yokota T., Wakasugi H., Yodoi J.;				
RT	"ATL-derived factor (ADF), an IL-2 receptor/Trac inducer homologous to				
RT	thioredoxin; possible involvement of dithiol-reduction in the IL-2				
RT	receptor induction.";				
RL	EMBO J. 8:757-764(1989).				
ON	[4]				
RP	SEQUENCE OF 1-14.				
RA	MEDLINE=91151337; PubMed=1998498;				
RT	Martin H., Dean M.;				
RL	"Identification of a thioredoxin-related protein associated with				
RL	plasma membranes.";				
RL	Biochem. Biophys. Res. Commun. 175:123-128(1991).				
ON	[5]				
RP	STRUCTURE BY NMR.				
RA	MEDLINE=90057393; PubMed=2684271;				
RT	Forman-Kay J.D., Clore G.M., Driscoll P.C., Wingfield P.,				
RL	Richards F.M., Gronenborn A.M.;				
RT	"A proton nuclear magnetic resonance assignment and secondary				
RT	structure determination of recombinant human thioredoxin.";				
RL	Biochemistry 28:7088-7097(1989).				
ON	[6]				
RP	STRUCTURE BY NMR.				
RA	MEDLINE=91159399; PubMed=2001356;				
RT	Forman-Kay J.D., Clore G.M., Wingfield P., Gronenborn A.M.;				
RL	"High-resolution three-dimensional structure of reduced recombinant				

RT human thiorodoxin in solution.";
 RL Biochemistry 30:2685-2698(1991).
 [7]
 RP STRUCTURE BY NMR.
 RX MEDLINE-95006318: PubMed-7922028:
 RA Qin J., Clore G.M., Gronenborn A.M.;
 RT "The high-resolution three-dimensional solution structures of the
 oxidized and reduced states of human thiorodoxin.";
 RL Structure 2:503-522(1994).
 [8]
 RP STRUCTURE BY NMR.
 RX MEDLINE-96347359: PubMed-8736558:
 RA Qin J., Clore G.M., Kennedy W.P., Kuszewski J., Gronenborn A.M.;
 RT "The solution structure of human thiorodoxin complexed with its
 target from Ref-1 reveals peptide chain reversal.";
 RL Structure 4:613-620(1996).
 [9]
 RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).
 RX MEDLINE-96399719: PubMed-8805557;
 RA Weichsel A., Gasdaska J.R., Powis G., Montfort W.R.;
 RT "Crystal structures of reduced, oxidized, and mutated human
 thiorodoxins: evidence for a regulatory homodimer.";
 RL Structure 4:735-751(1996).
 [10]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF MUTANT ASN-60.
 RX MEDLINE-96039128: PubMed-9369469;
 RA Andersen J.F., Sanders D.A., Gasdaska J.R., Weichsel A., Powis G.,
 Montfort W.R.;
 RT "Human thiorodoxin homodimers: regulation by pH, role of aspartate
 60, and crystal structure of the aspartate 60 --> asparagine
 mutant.";
 RL Biochemistry 36:13979-13988(1997).
 [11]
 RP ACTIVITY.
 RX MEDLINE-91097576: PubMed-2176490;
 RA Jacquot J.-P., de Lamotte F., Fontecave M., Schuermann P.,
 Decottignies P., Maginac-Maslow M., Wollman E.;
 RL "Human thiorodoxin reactivity-structure/function relationship.";
 CC Biochem. Biophys. Res. Commun. 173:1375-1381(1990).
 [12]
 RP FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
 THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL
 TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
 CC FUNCTION: ADF AUGMENTS THE EXPRESSION OF THE INTERLEUKIN-2
 RECEPTOR TAC (IL2R/P55).
 [13]
 RP SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

 DR EMBL: X77584: CA54687.1: -;
 DR EMBL: X54539: CA38410.1: -;
 DR EMBL: X54540: CA38410.1: JOINED.
 DR EMBL: X54541: CA38410.1: JOINED.
 DR EMBL: J04026: AAA74596.1: -;
 DR PIR: A31993: A31993.
 DR PIR: S04106: S04106.
 DR PIR: JH0568: JH0568.
 DR PIR: JH0568: JH0568.
 DR PDB: 3TRX: 15-JAN-93.
 DR PDB: 4TRX: 15-JAN-93.
 DR PDB: 1TRX: 30-SEP-94.
 DR PDB: 1TRU: 30-SEP-94.
 DR PDB: 1TRV: 30-SEP-94.
 DR PDB: 1TRW: 30-SEP-94.
 DR PDB: 1A10: 07-JUL-97.
 DR PDB: 1A10: 07-JUL-97.
 DR PDB: 1ERU: 01-AUG-96.
 DR PDB: 1ERU: 01-AUG-96.
 DR PDB: 1ERU: 14-OCT-96.
 DR PDB: 1ERU: 14-OCT-96.
 DR PDB: 1ERU: 14-OCT-96.

DR PDB: 1COG: 01-AUG-96.
 DR PDB: 1COH: 01-AUG-96.
 DR PDB: 1MDI: 03-JUN-95.
 DR PDB: 1MDJ: 20-JUL-95.
 DR PDB: 1MDK: 20-JUL-95.
 DR PDB: 1AUC: 25-FEB-98.
 DR PDB: 1AUC: 25-FEB-98.
 DR SWISS-2DPAGE: P10599: HUMAN.
 DR Swiss/Ghent-2DPAGE: 8006: IEF.
 DR PHCI-2DPAGE: P10599: -;
 DR Sienna-2DPAGE: P10599: -;
 DR MIM: 187700: -;
 DR InterPro: IPR000063: Thiorodoxin.
 DR Pfam: PF00085: Thiorod.
 DR PRINTS: PR00421: THIOREDOXIN.
 DR PROSITE: PS00194: THIOREDOXIN, 1.
 KW Redox-active center; Electron transport; 3D-structure.
 FT INT_MET 0
 FT DISULFID 31 34 REDOX-ACTIVE (BY SIMILARITY).
 FT CONFLICT 38 38 K -> N (IN REF. 2).
 FT CONFLICT 73 73 M -> T (IN REF. 2).
 FT STRAND 2 3
 FT HELIX 7 16
 FT TURN 18 19
 FT STRAND 22 27
 FT HELIX 32 35
 FT TURN 36 37
 FT TURN 38 41
 FT HELIX 42 43
 FT HELIX 44 47
 FT STRAND 52 57
 FT TURN 58 60
 FT HELIX 62 68
 FT TURN 69 69
 FT STRAND 75 80
 FT TURN 81 82
 FT STRAND 83 89
 FT HELIX 94 103
 SQ SEQUENCE 104 AA; 11606 MW; 7FFABDF3B6BE33A CRC64;
 Query Match 32.9%; Score 185; DB 1; Length 104;
 Best Local Similarity 35.7%; Pred. No. 3,3e-14;
 Matches 45; Conservative 11; Mismatches 28; Indels 42; Gaps 2;
 QY 2 VOIKIDNEFFETFLAAGHKILAVVQSSKRCGPKRMFPVFEHAE----- 47
 Db 1 VKQLESTAFQEAIDAGDKLVVDFAWCGPCMKIKPFHSLSEKYSNVIFLEVDVD 60
 QY 48 -----TCHITITPTFQMKFSOKVTLSFRIKRIICYSRSGFMNLIFFPGADAKKLEA 101
 Db 61 QQDVASECEVCMPTFFQFFKRGKRVG-----EFGANKKLEA 98
 QY 102 KTOELM 107
 Db 99 TINELV 104
 RESULT 2
 THIO_SHEEP
 ID THIO_SHEEP STANDARD; PRT; 104 AA.
 AC P50413;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thiorodoxin.
 GN TXN.
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 OX NCBI_TaxID=9940;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=95078463; PubMed=7987015;
 RA Droogmans L., Cleuter Y., Wollman E.E., Kettmann R., Burny A.;
 RT "Nucleotide sequence of ovine thioredoxin cDNA."
 RL DNA Seq. 4:277-279(1984).
 CC -1- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
 CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
 CC TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
 CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z25864; CAA81083.1; -.
 DR HSSP; P10599; IERT.
 DR InterPro; IPR000063; Thioredoxin.
 DR Pfam; PF00085; thiorcd; 1.
 DR PRINTS; PR00421; THIOREDOXIN.
 DR PROSITE; PS00194; THIOREDOXIN; 1.
 DR Redox-active center; Electron transport.
 FT INIT_MET 0 0 BY SIMILARITY.
 FT DISULFID 31 34 REDOX-ACTIVE (BY SIMILARITY).
 SO SEQUENCE 104 AA; 11711 MW; 506CE54C56A2208D CRC64;

Query Match 32.9%; Score 185; DB 1; Length 104;
 Best Local Similarity 34.9%; Pred. No. 3.3e-14;
 Matches 44; Conservative 13; Mismatches 27; Indels 42; Gaps 2;

OY 2 VQIKDNEKPTFLAGHKLAVQFSSKRCGPKRMFPVHELAE----- 47
 DB 1 VKQIESYVAQFQALNSAGELVYVDFSATWCGPKMKIPFHSLSEKXSNVVFLEVDVD 60
 OY 48 -----TCHIKTPTPFQMFKSKQVTLFSRIKRIICYSRSGFMSNLIFFPCGADAKKLEA 101
 DB 61 CQDVAAECVCKMPTFPQFKKGKVS-----EFGANKKLEA 98
 OY 102 KTOELM 107
 DB 99 TINELI 104

RESULT 3
 THIO_RABIT STANDARD; PRT; 104 AA.
 AC P08628;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thioredoxin.
 GN TXN.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Bone marrow;
 RX MEDLINE=88257078; PubMed=3164311;
 RA Johnson R.S., Mathews W.R., Biemann K., Hopper S.;
 RT "Amino acid sequence of thioredoxin isolated from rabbit bone marrow
 RT determined by tandem mass spectrometry."
 RL J. Biol. Chem. 263:9589-9597(1988).
 CC -1- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
 CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
 CC TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
 CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
 DR PIR; A28086; A28086.
 DR HSSP; P10599; IERV.

DR InterPro; IPR000063; Thioredoxin.
 DR Pfam; PF00085; thiorcd; 1.
 DR PRINTS; PR00421; THIOREDOXIN.
 DR PROSITE; PS00194; THIOREDOXIN; 1.
 DR Redox-active center; Electron transport.
 FT DISULFID 31 34 REDOX-ACTIVE (BY SIMILARITY).
 SO SEQUENCE 104 AA; 11629 MW; CAB6E5EBEC231F CRC64;

Query Match 32.7%; Score 184; DB 1; Length 104;
 Best Local Similarity 34.9%; Pred. No. 4.3e-14;
 Matches 44; Conservative 13; Mismatches 27; Indels 42; Gaps 2;

OY 2 VQIKDNEKPTFLAGHKLAVQFSSKRCGPKRMFPVHELAE----- 47
 DB 1 VKQIESYVAQFQALNSAGELVYVDFSATWCGPKMKIPFHSLSEKXSNVVFLEVDVD 60
 OY 48 -----TCHIKTPTPFQMFKSKQVTLFSRIKRIICYSRSGFMSNLIFFPCGADAKKLEA 101
 DB 61 CQDVAAECVCKMPTFPQFKKGKVS-----EFGANKKLEA 98
 OY 102 KTOELM 107
 DB 99 TINELI 104

RESULT 4
 THIO_BOVIN STANDARD; PRT; 104 AA.
 AC O97680;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thioredoxin.
 GN TXN.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gotoh S., Terashima H., Yagi K., Mizoguchi T.;
 RT "Cloning and sequence of bovine thioredoxin cDNA."
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
 CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
 CC TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
 CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF104105; AAC83380.1; -.
 DR HSSP; P10599; IERT.
 DR InterPro; IPR000063; Thioredoxin.
 DR Pfam; PF00085; thiorcd; 1.
 DR PRINTS; PR00421; THIOREDOXIN.
 DR PROSITE; PS00194; THIOREDOXIN; 1.
 DR Redox-active center; Electron transport.
 FT INIT_MET 0 0 BY SIMILARITY.
 FT DISULFID 31 34 REDOX-ACTIVE (BY SIMILARITY).
 SO SEQUENCE 104 AA; 11681 MW; 506CF9696A2208D CRC64;

Query Match 32.4%; Score 182; DB 1; Length 104;
 Best Local Similarity 34.9%; Pred. No. 7.3e-14;
 Matches 44; Conservative 12; Mismatches 28; Indels 42; Gaps 2;

```

OY      2 VOIDJTMKEFKFLTAAGHKLAVQFSSRCGCPCKRMFPVFEHLAE----- 47
       1 : : : | : ||| ||| : ||| : ||| :
DB      1 VKQISKAFOALMSAGEKLVVPFSATWCGPCMKIRPFHSLSSEKSINVFLEVDYD 60
OY      48 -----TCHIKTIPTFMQEKRSOKVTLESRIKRIRICCYRSGFMSMLIFECGADAKKLEA 101
       1 : : |||| ||| |||
DB      61 CODVAEEVEEVCMPPTFOFFKKQKG-----EFSGANKEKLEA 98
OY      102 KTOELM 107
       ||:
DB      99 TINELI 104

RESULT      5
THIO_MOUSE ID THIO_MOUSE STANDARD: PRT: 104 AA.
AC PI0639;
DT 01-JUN-1989 (Rel. 11, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thioresdoxin (ATL-derived factor) (ADF).
GN TXN.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89251607; PubMed=2785919;
RA Tagaya Y., Maeda Y., Mitsui A., Kndo N., Matsui H., Hamuro J.,
RA Brown N., Arai K.-I., Yokota T., Wakasugi H., Yodoi J.;
RT "Aut.-derived factor (ADF), an IL-2 receptor/Tac inducer homologous to
RT thioresdoxin; possible involvement of dihiol-reduction in the IL-2
RT receptor induction.";
RL EMBO J. 8:757-764(1989).
RN [2]
RP REVISIONS.
RX MEDLINE=94244626; PubMed=8187776;
RA Tagaya Y., Maeda Y., Mitsui A., Kndo N., Matsui H., Hamuro J.,
RA Brown N., Arai K.-I., Yokota T., Wakasugi H., Yodoi J.;
RL EMBO J. 13:2244-2244(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STAIN=129/SV; TISSUE=Liver;
RX MEDLINE=95137382; PubMed=7835695;
RA Matsui M., Taniguchi Y., Hirota K., Taketo M., Yodoi J.;
RT "Structure of the mouse thioresdoxin-encoding gene and its processed
RT pseudogene.";
RL Gene 152:165-171(1995).
CC -1- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL.
CC TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
CC -1- FUNCTION: ADF AUGMENTS THE EXPRESSION OF THE INTERLEUKIN-2
CC RECEPTOR TAC (IL2R/P55).
CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X77585; CAAS4688.1; -.
DR EMBL, D21859; BAA04881.1; -.
DR EMBL, D21855; BAA04881.1; JOINED.
DR EMBL, D21856; BAA04881.1; JOINED.
DR EMBL, D21857; BAA04881.1; JOINED.
DR EMBL, D21858; BAA04881.1; JOINED.
DR PTR, S04107; S04107.
RR HSSP; P10599; IERT.

```

```
DR SWISS-2DPAGE: P10639; MOUSE.  
DR MGI: 98874; Tn.  
DR InterPro: IPR000063; Thioredoxin.  
DR Pfam: PF00085; Thiored.1.  
DR PRINTS: PR00421; THIOREDOXIN.  
DR PROSITE: PS00194; THIOREDOXIN_1.  
DR Redox-active center; Electron transport.  
FT INIT MET 0  
FT DISULFID 31 34 REDOX-ACTIVE (BY SIMILARITY).  
SQ SEQUENCE 104 AA; 11544 MW; 60BE6196090AC773 CRC64;  
  
Query Match 32.0%; Score 180; DB 1; Length 104;  
Best Local Similarity 34.7%; Pred. No. 1,2e-13;  
Matches 43; Conservative 11; Mismatches 28; Indels 42; Gaps 2;  
  
QY 2 VOIKDNTNEFFTEFLTAACHKLAVVQFSKRCGPGCKRMFPVHEL----- 45  
|::: |::: |::: |::: |::: |::: |::: |::: |::: |:::  
DB 1 VKLIESKAEPALAAADKLAVVDFASTWCPCKMIPFHSLCDKYSNVVPLEVDVD 60  
|::: |::: |::: |::: |::: |::: |::: |::: |::: |:::  
QY 46 ----AETCHITPTPFQMEKRSOKVTLSRIKRICCRSGFMNSLIPEFGADAKKLEA 101  
|::: |::: |::: |::: |::: |::: |::: |::: |::: |:::  
DB 61 CQDVAADEVCMPTFGYYKKGGKVG-----EFGANKKEKLEA 98  
|::: |::: |::: |::: |::: |::: |::: |::: |::: |:::  
QY 102 KTOE 105  
|  
DB 99 SITE 102  
  
RESULT 6  
THIO_RAT STANDARD; PRT; 104 AA.  
ID THIO_RAT  
AC P11232;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Thioredoxin.  
GN TXN.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI_TaxID=10116;  
[1]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Liver;  
RX MEDLINE=89282399; PubMed=2734107;  
RA Tonissen K.F., Robins A.J., Wells J.R.E.;  
RT "Nucleotide sequence of a cDNA encoding rat thioredoxin.";  
RL Nucleic Acids Res. 17:3973-3973(1989).  
CC -!- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS  
THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL  
TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.  
CC -!- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outpost.  
CC The European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
DR EMBL: X14878; CA33019.1; .  
DR PIR: S04352; S04352.  
DR HSSP: P10599; IERT.  
DR InterPro: IPR000063; Thioredoxin.  
DR Pfam: PF00085; Thiored.1.  
DR PRINTS: PR00421; THIOREDOXIN.  
DR PROSITE: PS00194; THIOREDOXIN_1.  
DR Redox-active center; Electron transport.  
KW INT_MET 0  
KW DISULFID 31 34 REDOX-ACTIVE (BY SIMILARITY).  
FT INIT_MET 0  
SQ SEQUENCE 104 AA; 11542 MW; 60ACF196090ED773 CRC64;
```

```

QY      2 VOI|KDN|EKF|TLTAAGHKLAVVOFSSKRCGPCRMFPVHE|------ 45
       |:::| 1: |:::| 11111: |:::| 1: | 1
Db      1 VKLIESKEAFEDALAAADKLWVDVFSAATGCPCKMKIPFFSLDKYSNVFLEVDYD 60
QY      46 ----AECGHITPTPFQFGFKSOKVTLSRIKKITCCRRSGMSMLIFEGCADAKTLEA 101
       | 1: |:::| 11111: | 11111: | 11111: | 11111: | 11111: | 11111: |
Db      61 CQDYAADCEVCACMPTPFRKKGQKV-----EFSGANKEKLEA 98
QY      102 KTOE 105
       |
Db      99 TITE 102

RESULT   7
THIO_MACMU THIO_MACMU STANDARD; PRT; 104 AA.
AC          P29451;
DT          01-APR-1993 (Rel. 25, Created)
DT          01-APR-1993 (Rel. 25, Last sequence update)
DT          16-OCT-2001 (Rel. 40, Last annotation update)
DE          Thioedoxin.
GN          TXN.
OS          Macaca mulatta (Rhesus macaque).
OC          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC          Mammalia; Euarchia; Primates; Catarrhini; Cercopithecoidea;
OC          Cercopithecinae; Macaca.
OX          NCBI_TaxID=9544;
RN          [1]
RP          SEQUENCE FROM N.A.
RX          MEDLINE=92181438; PubMed=1543487;
RA          An G., Wu R.;
RT          "Thioedoxin gene expression is transcriptionally up-regulated by
RL          retinol in monkey conducting airway epithelial cells.";
RB          Biochem. Biophys. Res. Commun. 183:170-173(1992).
CC          -1- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
CC          THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
CC          TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
CC          -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC          -----
CC          This SWISS-PROT entry is copyright. It is produced through a collaboration
CC          between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC          the European Bioinformatics Institute. There are no restrictions on ways
CC          CC use by non-profit institutions as long as its content is in no way
CC          modified and this statement is not removed. Usage by and for commercial
CC          entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC          or send an email to license@isb-sib.ch).
CC          -----
Cc      EMBL; M84643; AAA36921.1; .
DR      PIR; JS0667; JS0667.
DR      HSSP; P10599; IERT.
DR      InterPro; IPRO00063; Thioedoxin.
DR      Pfam; PF00085; thioed: 1.
DR      PRINTS; PR00421; THIOREDOXIN.
DR      PROSITE; PS00194; THIOREDOXIN; 1.
KW      Redox-active center; Electron transport.
FT      INTMET 0 BY SIMILARITY.
FT      DISULFD 31 34 REDOX-ACTIVE (BY SIMILARITY).
SO      SEQUENCE 104 AA; 11606 MW; C804D5152FE8B70EB CRC64;

Query Match 31.9%; Score 179; DB 1; Length 104;
Best Local Similarity 34.7%; Pred. No. 1,6e-13;
Matches 43; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

2 VOI|KDN|EKF|TLTAAGHKLAVVOFSSKRCGPCRMFPVHE|------ 47
1 VKLIESKAFAFGALDAGUKLVVDFSAATGCPCKMKIPFFSLDSKYSNVFLEVYDD 60

```

QY	48	---	THITPTFEQFQKKSOKATLFSRIKRIITICRCYSGPMSNLIIEFGCADAKKLEA	101
Db	61	CODVASECEVVCMPFQCFKKQKYGK	-----EFGANKEKLEA	98
QY	102	KTOELM 107		
Db	99	TINELV 104		
RESULT	8			
THIO_CHICK	THIO_CHICK	STANDARD;	PRT; 104 AA.	
AC	P08629;			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-AUG-1988 (Rel. 08, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Thioredoxin.			
GN	TXN.			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88257080; PubMed=2838473;			
RA	Jones S.W., Luk K.-C.;			
RT	"Isolation of a chicken thioredoxin cDNA clone. Thioredoxin mRNA is			
RT	differentially expressed in normal and Rous sarcoma virus-transformed			
RT	chicken embryo fibroblasts."			
RL	J. Biol. Chem. 263:9607-9611(1988).			
CC	-1 FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS			
CC	THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,			
CC	TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.			
CC	-1 SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sdb.ch/announce/).			
CC	or send an email to license@isb-sdb.ch).			
CC	-----			
DR	EMBL: J03882; AAA49092.1; -.			
DR	PIR: A30006; A30006.			
DR	HSSP; P10599; 1ERT.			
DR	InterPro; IPR000063; Thioredoxin.			
DR	Pfam; PF00085; thioRedoxin.			
DR	PRINTS; PR00421; ThIOREDOXIN.			
DR	PROSITE; PS00194; THIOREDOXIN; 1.			
KW	Redox-active center; Electron transport.			
FT	INIT_MET 0			
FT	DISULFID 31 34			
SEQ	SEQUENCE 104 AA: 11569 MW: 60B6B5759010BB12 CRC64:			
		REDOX-ACTIVE (BY SIMILARITY).		
Query Match	28.5%;	Score 160;	DB 1;	Length 104;
Best local Similarity	31.0%;	Pred. No. 2.3e-11;		
Matches 39;	Conservative 15;	Mismatches 30;	Indels 42;	Gaps 2;
QY	2	VOIIDTNEEFKFTLAAGHKILAVOFSKRCGPKRMFPVFHEL	-----	45
Db	1	VKSVGNLADFEELKLAGKELVVDPSATWCGPKMKIKRFFHSLDCKRGDVFYIELDVD	60	
QY	46	---AETGHIKITIPFPQFKKSQKVTLSRIKRIITICRCYSGPMSNLIIEFGCADAKKLEA	101	
Db	61	AADVATHDVCVKMPFQCFKKQKYGK	-----EFGANKEKLEA	98
QY	102	KTOELM 107		
Db	99	TIKSLV 104		

```

RESULT 9
THIO_COPCM STANDARD: PRT: 106 AA.
ID THIO_COPCM
AC 09UM02:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ThioRedoxin (Allergen Cop c 2).
OS Coprinus comatus (Shaggy mane).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Coprinaceae; Coprinus.
OX NCBI_TaxID=56187;
RN [1]
RP SEQUENCE FROM N.A.
RA Brander K.A., Cramer R., Schuermann P., Pichler W.J., Helbling A.;
RT "Coprinus thioRedoxin as inhalative allergen and crossreactive human
RT autoantigen."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
CC TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ242791; CAB52130.1; -.
DR HSSP: P80028; 1TQF.
DR InterPro: IPR000063; ThioRedoxin.
DR Pfam: PF00085; ThioRed. 1.
DR PRINTS: PR00421; THIOREDOXIN.
DR PROSITE: PS00194; THIOREDOXIN; 1.
RW Redox-active center; Electron transport; Allergen.
FT DISULFID 30 33 REDOX-ACTIVE (BY SIMILARITY).
FT SEQUENCE 106 AA; 11772 MW; 05A2155B210B8C69 CRC64;
SQ

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98139906; PubMed=9473519;
RA Miranda-Vizuet A., Gustafsson J.-A., Spyrou G.;
RT "Molecular cloning and expression of a cDNA encoding a human
RT thioRedoxin-like protein."
RL Biochem. Biophys. Res. Commun. 243:284-288(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Lee K.-K., Murakawa M., Takahashi S., Tsubuki S., Sakamaki K.,
RA Yonehara S.;
RT "Purification, cloning and characterization of TRP32, a novel
RT thioRedoxin-related protein of 32 kDa."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Zhou Y., Pan M.H., Yuan J.G., Qiang B.Q.;
RT "The discovery of a new gene that has high homology to the human
RT thioRedoxin gene."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20284953; PubMed=10826702;
RA Miranda-Vizuet A., Spyrou G.;
RT "Genomic structure and chromosomal localization of human
RT thioRedoxin-like protein gene (txl).";
RL DNA seq. 10:419-424(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.22 ANGSTROMS) OF 1-108.
RA Jin J., Chen X., Guo Q., Yan J., Qiang B., Rao Z.;
RT "Crystal structure of the catalytic domain of a new human thioRedoxin-
RT like protein."
RL Submitted (NOV-2000) to the PDB data bank.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF003938; AAC39599.1; -.
DR EMBL: AF052659; AAC39898.1; -.
DR EMBL: AF051896; AAC05830.1; -.
DR EMBL: AF143897; AAF66676.1; -.
DR EMBL: AF143897; AAF66676.1; JOINED.
DR EMBL: AF143890; AAF66676.1; JOINED.
DR EMBL: AF143891; AAF66676.1; JOINED.
DR EMBL: AF143892; AAF66676.1; JOINED.
DR EMBL: AF143893; AAF66676.1; JOINED.
DR EMBL: AF143894; AAF66676.1; JOINED.
DR EMBL: AF143895; AAF66676.1; JOINED.
DR EMBL: AF143896; AAF66676.1; JOINED.
DR EMBL: BC001156; AAH01156.1; -.
DR PDB: 1GH2; 01-MAY-01.
DR MIM: 603049; -.
DR InterPro: IPR000063; ThioRedoxin.
DR Pfam: PF00085; ThioRed. 1.
DR PROSITE: PS00194; THIOREDOXIN; 1.
RW Redox-active center; Electron transport; 3D-structure.
FT DOMAIN 1 109 THIOREDOXIN.
FT DISULFID 34 37 REDOX-ACTIVE.
FT SEQUENCE 289 AA; 32251 MW; B2CC0BD8042225C2 CRC64;
SQ

```



```

DR   PCHI-2DPAGE; Q9Z7P5; -.
DR   TIGR: CP0088; -.
DR   InterPro: IPR000063; Thioresdoxin.
DR   Pfam: PF00085; Thioresdoxin.
DR   PRINTS: PR00421; THIOREDOXIN.
DR   PROSITE: PS00194; THIOREDOXIN_1.
KW   Redox-active center; Electron transport; Complete proteome.
FT   DISULFID 28      REDOX-ACTIVE (BY SIMILARITY).
SQ   SEQUENCE 102 AA; 11292 MW; 76A190218324BA68 CAC64;

Query Match          19.8%; Score 111; DB 1; Length 102;
Best Local Similarity 29.6%; Pred. No. 8,1e-06;
Matches 32; Conservative 18; Mismatches 28; Indels 30; Gaps 5;

OY   1 MVGIKDTNEKEFLTLTAAGHKLAVVQFSKKRGPCRMFPVFHEI----- 45
    ||||| | : | : | : | : | : | : | : | : | : | : | : | : |
DB   1 MKVIISSEN-PDSFI-ASG--LVLDPFPAEWCGPCRMILTPILENTAAELPHVTIKINID 56
    ||||| | : | : | : | : | : | : | : | : | : | : | : | : |

OY   46 -----AEFGCHKITPTFMFKKSOKVLTLSFRKRIRICCRSPFNMLI 88
    ||| : ||| : || : | : | : | : | : | : | : | : | : | : |
DB   57 ENSKPATYEVSISPTLIFRDGNVA-----RVVGLKDKEEFLTNLI 98
    ||| : ||| : || : | : | : | : | : | : | : | : | : | : |

RESULT 12
TRX2_SCHPO STANDARD; PROT; 102 AA.
AC O14463; Q9UTS9;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thioredoxin II (TR-II).
GN TRX2 OR SPACTD4.07C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_Taxid=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21167391; PubMed=11267679;
RA Cho Y.-W., Shin Y.H., Kim Y.-T., Kim H.-G., Lee Y.-J., Park E.-H.,
RA Fuchs J.A., Lim C.-U.;
RT "Characterization and regulation of Schizosaccharomyces pombe gene
RT encoding thioredoxin.";
RL Blochim. Biophys. Acta 1518:194-199(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Lenaers G., Perret E., Bonin O., Picard A., Caput D.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Gentiles S., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
CC TO A DISULFIDE & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
CC -!- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce1a)
CC or send an email to license@isb-sib.ch).
CC -----
DR   EMBL: AF251279; AAF76881.1; -
DR   EMBL: AJ0003819; CA060033.1; -
DR   EMBL: Z995332; CAB16724.1; -.
DR   HSSP: P80028; 110F.
```

DR InterPro: IPR000063; ThioRedoxin.
 DR Pfam: PF00085; ThioRedoxin.
 DR PRINTS: PR00421; THIOREDOXIN.
 DR PROSITE: PS00194; THIOREDOXIN; 1.
 KW Redox-active center; Electron transport.
 FT INTL MET 0 BY SIMILARITY.
 FT DISULFID 29 32 REDOX-ACTIVE (BY SIMILARITY).
 SO SEQUENCE 102 AA; 11166 MW; 7069F4ACDAC34595 CRC64;

Query Match 19.6%; Score 110; DB 1; Length 102;
 Best Local Similarity 25.0%; Pred. No. 1e-05;
 Matches 30; Conservative 17; Mismatches 29; Indels 44; Gaps 3;

QY 2 VOIINDNEFFETLTAGHKLAVQSSKRCGPCRMPPEVH----- 43
 ID THIF_PEA STANDARD; PRT; 102 AA.
 AC P29450;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ThioRedoxin I (TR-1).
 GN TRX1.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 NX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21167391; PubMed=11267679;
 RA Cho Y.-W., Shin Y.H., Kim Y.-T., Kim H.-G., Lee Y.-J., Park E.-H.,
 RA Fuchs J.A., Lim C.-J.;
 RT "Characterization and regulation of Schizosaccharomyces pombe gene
 RT encoding thioRedoxin.";
 RL Biochim Biophys Acta 1518:194-199(2001).
 CC -I- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
 CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
 CC TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
 CC -I- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF192765; AAF05765.1; -
 DR InterPro: IPR000063; ThioRedoxin.
 DR Pfam: PF00085; ThioRedoxin.
 DR PRINTS: PR00421; THIOREDOXIN.
 DR PROSITE: PS00194; THIOREDOXIN; 1.
 KW Redox-active center; Electron transport.
 FT INTL MET 0 BY SIMILARITY.
 FT DISULFID 29 32 REDOX-ACTIVE (BY SIMILARITY).
 SO SEQUENCE 102 AA; 11307 MW; 70781D7BADB2EAB5 CRC64;

Query Match 19.2%; Score 108; DB 1; Length 102;
 Best Local Similarity 24.0%; Pred. No. 1.8e-05;
 Matches 29; Conservative 18; Mismatches 30; Indels 44; Gaps 3;

QY 2 VOIINDNEFFETLTAGHKLAVQSSKRCGPCRMPPEVH----- 43
 ID THIF_PEA STANDARD; PRT; 182 AA.
 AC P29450;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ThioRedoxin F-type, chloroplast precursor (TRX-F).
 DE Pisum sativum (Garden pea).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
 NX NCBI_TaxID=3888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96215867; PubMed=1581563;
 RA Lepiniec L., Hodges M., Gadal P., Cretin C.;
 RT "Isolation, characterization and nucleotide sequence of a full-length
 RT pea cDNA encoding thioRedoxin-F";
 RL Plant Mol. Biol. 18:1023-1025(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96215867; PubMed=8642611;
 RA Sahravay M., Hecht V., Lopez Jaramillo J., Chueca A., Chartier Y.,
 RA Meyer Y.;
 RT "Iron position as an evolutionary marker of thioRedoxins and
 RT thioRedoxin domains.";
 RL J. Mol. Evol. 42:422-431(1996).
 CC -I- FUNCTION: PARTICIPATES IN VARIOUS REDOX REACTIONS THROUGH THE
 CC REVERSIBLE OXIDATION OF THE ACTIVE CENTER DITHIOL, TO A DISULFIDE.
 CC THE F FORM IS KNOWN TO ACTIVATE A NUMBER OF ENZYMES OF THE
 CC PHOTOSYNTHETIC CARBON CYCLE.
 CC -I- SUBCELLULAR LOCATION: Chloroplast.
 CC -I- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY. PLANT F-TYPE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X63537; CAA45098.1; -
 DR EMBL: U35830; AAC49357.1; -
 DR PIR: S20929; S20929.
 DR HSSP: P10599; IAIU
 DR InterPro: IPR000063; ThioRedoxin.
 DR Pfam: PF00085; ThioRedoxin.
 DR PROSITE: PS00194; THIOREDOXIN; 1.
 KW Redox-active center; Electron transport; Chloroplast; Transit peptide.
 FT TRANSIT 1 69 CHLOROPLAST (POTENTIAL).
 FT CHAIN 70 182 THIOREDOXIN F-TYPE.
 FT DISULFID 106 109 REDOX-ACTIVE (BY SIMILARITY).
 SO SEQUENCE 182 AA; 19775 MW; 158FC352CB950F1 CRC64;

Query Match 19.0%; Score 106.5; DB 1; Length 182;
 Best Local Similarity 29.9%; Pred. No. 4.6e-05;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2002, 03:25:56 ; Search time 48.76 Seconds

(without alignments)
210.861 Million cell updates/sec

Title: US-09-746-783-19

Perfect score: 562

Sequence: 1 MVLKIDPNEKFTFLTAGH.....IFECGADAKKLEAKTQELM 107

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : PIR_71:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	190	33.8	105	1 JH0568	thioredoxin [valid
2	185	32.9	105	1 S04107	thioredoxin - mus
3	184	32.7	104	1 A28086	thioredoxin - rabb
4	184	32.7	105	1 S04352	thioredoxin - rat
5	182	32.4	105	1 JS0667	thioredoxin - rhes
6	165	29.4	105	1 A30006	thioredoxin - chic
7	118	21.0	289	2 JCS938	thioredoxin-like p
8	115	20.5	103	2 T339085	thioredoxin II - f
9	111	19.8	102	2 D72052	thioredoxin CP0088
10	111	19.8	102	2 H86572	thioredoxin f prec
11	106.5	19.0	182	2 S20929	thioredoxin f prec
12	103.5	18.4	190	2 S04651	thioredoxin f prec
13	103	18.3	148	2 B96721	probable thioredox
14	102.5	18.2	104	1 TXB11	thioredoxin I - ye
15	101.5	18.1	107	1 S47867	thioredoxin-like p
16	100.5	17.9	133	2 S58123	thioredoxin (clone
17	100.5	17.9	182	2 T07837	thioredoxin f prec
18	99	17.6	105	1 A46264	thioredoxin 1 - sl
19	98	17.4	104	2 C46264	thioredoxin 3 - sl
20	96.5	17.2	191	2 T12261	thioredoxin f prec
21	95.5	17.0	121	2 T33987	thioredoxin II - f
22	95	16.9	102	2 C81660	thioredoxin 11 - f
23	94	16.7	102	2 B71503	probable thioredox
24	94	16.7	131	2 T50865	thioredoxin-like p
25	94	16.7	131	2 T50864	thioredoxin-like p
26	94	16.7	131	2 T50862	thioredoxin-like p
27	94	16.7	281	2 S49353	protein S2 - Pala
28	94	16.7	282	2 S49352	protein S1 - Pala
29	93.5	16.6	275	2 T00710	thioredoxin homolo

30	92.5	16.5	113	1 S57775	thioredoxin h, cyt
31	91	16.2	221	2 T04271	probable thioredox
32	90.5	16.1	122	2 T04090	thioredoxin-like p
33	89	15.8	88	2 B46264	thioredoxin 2 - sl
34	89	15.8	103	1 TXB12	thioredoxin II - y
35	88.5	15.7	107	1 S31915	thioredoxin - red
36	88.5	15.7	107	2 T33843	hypothetical prote
37	88.5	15.7	131	2 T50863	thioredoxin-like p
38	86	15.3	118	2 S58120	thioredoxin (clone
39	85.5	15.2	106	2 A49888	thioredoxin - Peni
40	85.5	15.2	114	1 J02242	thioredoxin h - Ar
41	85.5	15.2	172	1 S38909	thioredoxin m prec
42	85.5	15.2	302	2 A96789	protein T23E18.2 f
43	85	15.1	119	2 T50867	thioredoxin-like p
44	84.5	15.0	156	2 T08086	tyrosine light chain
45	84.5	15.0	236	2 T09930	thioredoxin homolo

ALIGNMENTS

RESULT 1

JH0568

thioredoxin [validated] - human

N:Alternate names: ATL-derived factor (ADF); eosinophil cytotoxicity-enhancing factor

C:Species: Homo sapiens (man)

C>Date: 30-Jun-1992 #sequence.revision 26-May-1994 #text.change 15-Sep-2000

C:Accession: JH0568; S04106; S44375; A31993; P10079; A60749; A38922; S53453; A60870

R:Tagaya, Y.; Maeda, Y.; Mitsui, A.; Kondo, N.; Matsui, H.; Hamuro, J.; Brown, N.; Al

Gene 102, 221-228, 1991

A:Title: Isolation and characterization of human thioredoxin-encoding genes.

A:Reference number: JH0568; MUID:91340156

A:Molecule type: DNA

A:Residues: 1-105 <TAG1>

A:Cross-references: EMBL:X54539; NID:937455; PIDN:CAA38410.1; PID:9825724; EMBL:X5454

R:Tagaya, Y.; Maeda, Y.; Mitsui, A.; Kondo, N.; Matsui, H.; Hamuro, J.; Brown, N.; Al

EMBO J. 8, 757-764, 1989

A:Title: ATL-derived factor (ADF), an IL-2 receptor/Tac inducer homologous to thiore

A:Reference number: S04106; MUID:99231607

A:Molecule type: mRNA

A:Accession: S04106

A:Residues: 1-105 <TAG1>

A:Cross-references: GB:X77584; NID:9453963; PIDN:CAA54687.1; PID:9453964

A>Note: this sequence has been revised in reference S44375

EMBO J. 13, 2244, 1994

A:Reference number: S44375; MUID:94244626

A:Contents: erratum

A:Accession: S44375

A:Molecule type: mRNA

A:Residues: 1-105 <TAG2>

A:Cross-references: EMBL:X77584; NID:9453963; PIDN:CAA54687.1; PID:9453964

R:Mollman, E.E.; d'Aurilol, L.; Rimsky, L.; Shaw, A.; Jacquot, J.P.; Wingfield, P.; G

J. Biol. Chem. 263, 15506-15512, 1988

A:Title: Cloning and expression of a cDNA for human thioredoxin.

A:Reference number: A31993; MUID:89008454

A:Molecule type: mRNA

A:Accession: A31993

A:Residues: 1-38, 'N', '40-73', '75-105 <WOL>

A:Cross-references: GB:J04026; NID:9339648; PIDN:AAA74596.1; PID:9339649

R:Martin, H.; Dean, M.

Biochem. Biophys. Res. Commun. 175, 123-128, 1991

A:Title: Identification of a thioredoxin-related protein associated with plasma memb

A:Reference number: P10079; MUID:91151337

A:Molecule type: protein

A:Residues: 2-13, 'X', '15 <MAR>

R:Silberstein, D.S.; Ali, M.H.; Baker, S.L.; David, J.R.

J. Immunol. 143, 979-983, 1989

A:Title: Human eosinophil cytotoxicity-enhancing factor. Purification, physical char

A:Reference number: A60749; MUID:89309777

A:Accession: A60749

A:Molecule type: protein
A:Residues: 2-12, 'K', '14-15, 'XX', '18-19, 'X', '21-22 <STL>
A:Note: The abstract is inconsistent with figure 4 in having one undetermined residue after
R:Rinsky, L., Waksugi, H., Ferrate, P., Robin, P., Capdevielle, J., Tursz, T., Fradelizi
J. Immunol. 136, 3304-3310, 1986
A:Title: Purification to homogeneity and NH-2-terminal amino acid sequence of a novel induc
A:Reference number: A38922; MUID:86169684
A:Accession: A38922
A:Molecule type: protein
A:Residues: 2-16 <WAK>
R:Dean, M.F., Martin, H., Sansom, P.A.
Biochem. J. 304, 861-867, 1994
A:Title: Characterization of a thioredoxin-related surface protein.
A:Reference number: S53453; MUID:95118305
A:Accession: S53453
A:Molecule type: protein
A:Residues: 1-21;38-57 <DEA>
A:Note: described to be a surface-associated thioredoxin
R:Waksugi, H., Rinsky, L., Mahe, Y., Kamel, A.M., Fradelizi, D., Tursz, T., Bertoglio,
Proc. Natl. Acad. Sci. U.S.A. 84, 804-808, 1987
A:Title: Epstein-Barr virus-containing B-cell line produces an interleukin 1 that it uses
A:Reference number: A60870; MUID:87118252
A:Contents: annotation
R:Weichsel, A., Gasdaska, J.R., Powis, G., Montfort, W.R.
submitted to the Brookhaven Protein Data Bank, February 1996
A:Reference number: A65533; PDB:1ERT
A:Contents: annotation; X-ray crystallography, 1.7 angstroms, reduced form, residues 1-1
R:Weichsel, A., Gasdaska, J.R., Powis, G., Montfort, W.R.
submitted to the Brookhaven Protein Data Bank, February 1996
A:Reference number: A65534; PDB:1ERU
A:Contents: annotation; X-ray crystallography, 2.1 angstroms, oxidized form, residues 1-1
R:Forman-Kay, J.D., Clore, G.M., Gronenborn, A.M.
submitted to the Brookhaven Protein Data Bank, December 1990
A:Reference number: A50924; PDB:4TRX
A:Contents: annotation; conformation by (1)H-NMR, residues 1-73, 'T', '75-105
R:Forman-Kay, J.D., Clore, G.M., Wingfield, P.T., Gronenborn, A.M.
Biochemistry 30, 2685-2698, 1991
A:Title: High-resolution three-dimensional structure of reduced recombinant human thior
A:Reference number: A38953; MUID:91159399
A:Contents: annotation; conformation by (1)H- and (15)N-NMR
A:Comment: This small ubiquitous protein functions in many intracellular biological path
C:Genetics:
A:Gene: GDB:TXN
A:Cross-references: GDB:120475; OMIM:187700
A:Map position: 9q31-9q31
A:Introns: 8/3; 43/3; 63/3; 85/3
C:Superfamily: thioredoxin; thioredoxin homology
C:Keywords: redox-active disulfide
E:2-105/Product: thioredoxin #status experimental <MAT>
E:9-92/Domains: thioredoxin homology <THR>
E:32-35/Disulfide bonds: redox-active #status experimental

```

Query Match          33.8%; Score 190; DB 1; Length 105;
Best Local Similarity 36.2%; Pred. No. 1.3e-13;
Matches 46; Conservative 11; Mismatches 26; Indels 42; Gaps 2;

QY      1 MWQIKIDNEKFTPLAAGHKLVAVQFSSKRCPCKRMFPVPHELAE----- 47
       || : | : | | | | | | : | | | : | | | : |
Db      1 MKVQESTATPGEALDAAGDKLVVDPSATWCGPKMKLKFPHSLSEKSNVFLEVDVD 60
QY      48 -----TCHIKTIPTPFOMFKSKOKVTLSFRILRIICCRSGSMNLIFPCGADAKKLE 100
       | : | : | | | | | | | | | | | | | | | | | | | | | |
Db      61 DCQDVASCEVEYKCMPTFOFFKKGRKV-----EFGANKKEKLE 98
QY      101 AKTOELM 107
       | | | |
Db      99 ATINELY 105

RESULT      2
S04107
thioredoxin - mouse
```

N:Alternate names: ATL-derived factor (ADF)
C:Species: Mus musculus (house mouse)
C:Date: 21-Nov-1993 #sequence_revision 17-Oct-1997 #text_change 11-Jun-1999
C:Accession: J04068; S44376; S04107
R:Matsui, M.; Taniguchi, Y.; Hirota, K.; Taketo, M.; Yodai, J.
Gene 152, 165-171, 1995
A:Title: Structure of the mouse thioredoxin-encoding gene and its processed pseudogen
A:Reference number: J04068; MUID:95137382
A:Accession: J04068
A:Molecule type: DNA
A:Residues: 1-105 <MAT>
A:Cross-references: DDBJ:D21855; NID:9517128
R:Tataya, Y.; Maeda, Y.; Mitsu, A.; Kando, N.; Matsui, H.; Hamuro, J.; Brown, N.; Ar
EMBO J. 13, 2244, 1994
A:Reference number: S44375; MUID:94244626
A:Contents: erratum
A:Accession: S44376
A:Molecule type: mRNA
A:Residues: 1-105 <TG1>
A:Cross-references: EMBL:X7585; NID:9453971; PIDN:CAA54688.1; PID:9453972
R:Tataya, Y.; Maeda, Y.; Mitsu, A.; Kondo, N.; Matsui, H.; Hamuro, J.; Brown, N.; Ar
EMBO J. 8, 757-764, 1989
A:Title: ATL-derived factor (ADF), an IL-2 receptor/Tac inducer homologous to thiore
A:Reference number: S04106; MUID:89251607
A:Accession: S04107
A:Molecule type: mRNA
A:Residues: 1-93, 'N', '94-96', 'ALT', '100-104', 'S' <TAG2>
A:Cross-references: GB:X77585
C:Comment: This small ubiquitous protein functions in many intracellular biological p
C:Genetics:
A:Gene: MGI:Txn
A:Cross-references: MGI:36258
A:Map position: 4:24.6
A:Introns: 29/2; 44/1; 84/2
C:Superfamily: thioredoxin; thioredoxin homology
C:Keywords: redox-active disulfide
E:9-32/Domain: thioredoxin homology <THR>
E:32-35/Disulfide bonds: redox-active #status predicted

```

Query Match Summary
Score: 35.2%
Pred. No.: 4
Length: 105
Matches: 44
Conservative: 11
Mismatches: 28
Indels: 42
Gaps: 2

Query: A28086
Subject: A28086
Species: Oryctolagus cuniculus (domestic rabbit)
C.Date: 30-Jun-1989
#sequence_revision: 26-May-1994
#text_change: 19-Oct-1995
R.Johnson, R.S.; Mathews, W.R.; Biemann, K.; Hopper, S.
J. Biol. Chem. 263, 9589-9597, 1988
A.Title: Amino acid sequence of thioredoxin isolated from rabbit bone marrow
A.Reference number: A28086; MUID:88257078
A.Accession: A28086
A.Molecule type: protein
A.Residues: 1-104
<JOB>
C.Keywords: redox-active disulfide
C.Superfamily: thioredoxin; thioredoxin homology
C.Domain: thioredoxin homology
<THR>

RESULT 3
A28086
thioredoxin - rabbit
C.Species: Oryctolagus cuniculus (domestic rabbit)
C.Date: 30-Jun-1989
#sequence_revision: 26-May-1994
#text_change: 19-Oct-1995
C.Accession: A28086
R.Johnson, R.S.; Mathews, W.R.; Biemann, K.; Hopper, S.
J. Biol. Chem. 263, 9589-9597, 1988
A.Title: Amino acid sequence of thioredoxin isolated from rabbit bone marrow
A.Reference number: A28086; MUID:88257078
A.Accession: A28086
A.Molecule type: protein
A.Residues: 1-104
<JOB>
C.Keywords: redox-active disulfide
C.Superfamily: thioredoxin; thioredoxin homology
C.Domain: thioredoxin homology
<THR>

```

F;31-34/Disulfide bonds: redox-active #status predicted

Query Match

32.7%; Score 184; DB 1; Length 104;

Best Local Similarity 34.9%; Pred. No. 5.7e-13;

Matches 44; Conservative 13; Mismatches 27; Indels 42; Gaps 2;

QY

2 VQIITKPTNEKFTLTAAGHKLAVVQFSKRCGPKRMFPVPHLEA----- 47

Db

1 VKQIESKAFOEVLDSAGDKLVVDFSATWCGCKMKRPFHSLSEKFNWVLEVDVD 60

QY

48 -----TCHIKITPTFQMFKRSOKVTLFSRIKRIICCYRSGFMSNLIFFECGADAKKLEA 101

Db

61 CKDIAACEVCKMPTFOFFKKGKVG-----EFGANKKEKLEA 98

QY

102 KTOELM 107

Db

99 TINELL 104

RESULT 4

S04352

thioredoxin - rat

N:Alternate names: thioredoxin-related surface protein SASP

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Mar-1990 #sequence_revision 26-May-1994 #text_change 11-Jun-1999

C:Accession: S04352; S66372

R:Tonnissen, K.F.; Robins, A.J.; Wells, J.R.E.

Nucleic Acids Res. 17, 3973, 1989

A:Title: Nucleotide sequence of a cDNA encoding rat thioredoxin.

A:Reference number: S04352; MUID:89282399

A:Accession: S04352

A:Molecule type: mRNA

A:Residues: 1-105 <TRON>

A:Cross-references: EMBL:X14878; NID:g57385; PIDN:CAA33019.1; PID:g57386

R:Dean, M.F.; Martin, H.; Sansom, P.A.

Biochem. J. 304, 861-867, 1994

A:Title: Characterization of a thioredoxin-related surface protein.

A:Reference number: S53453; MUID:95118305

A:Accession: S66372

A:Molecule type: protein

A:Residues: 2-21 <DEA>

A:Note: 12-Lys, 15-Gly, 16-Leu, 17-Gln, and 18-Leu were also found

C:Superfamily: thioredoxin; thioredoxin homology

C:Keywords: redox-active disulfide

F;2-105/Product: thioredoxin #status experimental <MAT>

F;9-92/Domain: thioredoxin homology <THR>

F;32-35/Disulfide bonds: redox-active #status predicted

Query Match

32.7%; Score 184; DB 1; Length 105;

Best Local Similarity 35.2%; Pred. No. 5.7e-13;

Matches 44; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

QY

1 MVOIITKPTNEKFTLTAAGHKLAVVQFSKRCGPKRMFPVPHLEA----- 45

Db

1 MVLIESKEAFQELAAAGKLVVDFSATWCGCKMKRPFHSLCDKYSNVVLEVDVD 60

QY

46 -----AETCHIKITPTFQMFKRSOKVTLFSRIKRIICCYRSGFMSNLIFFECGADAKKLE 100

Db

61 DCQDVADCEVCKMPTFOFFKKGKVG-----EFGANKKEKLE 98

QY

101 AKTOEL 105

Db

99 ATITE 103

RESULT 5

JS0667

thioredoxin - rhesus macaque

C:Species: Macaca mulatta (rhesus macaque)

C:Date: 30-Jun-1992 #sequence_revision 26-May-1994 #text_change 11-Jun-1999

C:Accession: JS0667

R:An, G.; Wu, R.

Biochem. Biophys. Res. Commun. 183, 170-175, 1992

A:Title: Thioredoxin gene expression is transcriptionally up-regulated by retinol in

A:Reference number: JS0667; MUID:92181438

A:Accession: JS0667

A:Molecule type: mRNA

A:Residues: 1-105 <RANG>

A:Cross-references: GB:M84643; NID:g342338; PIDN:AAA36921.1; PID:g342339

C:Superfamily: thioredoxin; thioredoxin homology

C:Keywords: redox-active disulfide

F;9-92/Domain: thioredoxin homology <THR>

F;32-35/Disulfide bonds: redox-active #status predicted

Query Match

32.4%; Score 182; DB 1; Length 105;

Best Local Similarity 35.4%; Pred. No. 9.5e-13;

Matches 45; Conservative 11; Mismatches 29; Indels 42; Gaps 2;

QY

1 MVOIITKPTNEKFTLTAAGHKLAVVQFSKRCGPKRMFPVPHLEA----- 47

Db

1 MVKQIESKAFOEVLDSAGDKLVVDFSATWCGCKMKRPFHSLSEKYSNVVLEVDVD 60

QY

48 -----TCHIKITPTFQMFKRSOKVTLFSRIKRIICCYRSGFMSNLIFFECGADAKKLE 100

Db

61 DCQDVADCEVCKMPTFOFFKKGKVG-----EFGANKKEKLE 98

QY

101 AKTOELM 107

Db

99 ATINELV 105

RESULT 6

A30006

thioredoxin - chicken

C:Species: Gallus gallus (chicken)

C:Date: 31-Mar-1989 #sequence_revision 26-May-1994 #text_change 11-Jun-1999

C:Accession: A30006

R:Jones, S.W.; Luk, K.C.

J. Biol. Chem. 263, 9607-9611, 1988

A:Title: Isolation of a chicken thioredoxin cDNA clone: thioredoxin mRNA is different

A:Reference number: A30006; MUID:88257080

A:Accession: A30006

A:Molecule type: mRNA

A:Residues: 1-105 <ION>

A:Cross-references: GB:J03882; NID:g212765; PIDN:AAA49092.1; PID:g212766

C:Superfamily: thioredoxin; thioredoxin homology

C:Keywords: redox-active disulfide

F;9-92/Domain: thioredoxin homology <THR>

F;32-35/Disulfide bonds: redox-active #status predicted

Query Match

29.4%; Score 165; DB 1; Length 105;

Best Local Similarity 31.5%; Pred. No. 6.9e-11;

Matches 40; Conservative 15; Mismatches 30; Indels 42; Gaps 2;

QY

1 MVOIITKPTNEKFTLTAAGHKLAVVQFSKRCGPKRMFPVPHLEA----- 45

Db

1 MWSVGNLADFEALKAAGKLVVDFSATWCGCKMKRPFHSLCDKFGDVVLEVDVD 60

QY

46 -----AETCHIKITPTFQMFKRSOKVTLFSRIKRIICCYRSGFMSNLIFFECGADAKKLE 100

Db

61 DAQDVATHQVCKMPTFOFFKKGKVG-----EFGANKKEKLE 98

QY

101 AKTOELM 107

Db

99 ETIKSLV 105

RESULT 7

JC3938

thioredoxin-like protein - human

C:Species: Homo sapiens (man)


```

Db 1 MWKLISEN-FDSFT-ASG--LVLVDFEAWCGPCRMFLPTELELAELPHVTIGKINID 56
OY 46 -----AETCHIKITPTFQMFKKSOKVTLFSRIKRIICCYSGENSLI 88
Db 57 ENSKPAETVESSIPFLILFRDGENVA-----RVGLKDKKEFLTNLI 98

RESULT 11
S20929
Chloredoxin f precursor - garden pea
C:Species: Pisum sativum (garden pea)
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 11-Jun-1999
C:Accession: S20929
R:Leptiec, L.; Hodges, M.; Gadal, P.; Cretin, C.
Plant Mol. Biol. 18, 1023-1025, 1992
A:Title: Isolation, characterization and nucleotide sequence of a full-length pea cDNA
A:Reference number: S20929; MUID:92256804
A:Accession: S20929
A:Molecule type: mRNA
A:Residues: 1-182 <LEP>
A:Cross-references: EMBL:X63537; NID:g20906; PIDN:CAA45098.1; PID:g20907
C:Superfamily: Thioedoxin; thioedoxin homology
C:Keywords: chloroplast; redox-active disulfide
F:1-74/Domain: transit peptide (chloroplast) #status predicted <TNP>
F:75-182/Product: thioedoxin f #status predicted <MAT>
F:83-167/Domain: thioedoxin homology <THR>
F:106-109/Disulfide bonds: redox-active #status predicted

Query Match 19.0%; Score 106.5; DB 2; Length 182;
Best Local Similarity 29.9%; Pred. No. 0.00029;
Matches 26; Conservative 17; Mismatches 21; Indels 23; Gaps 3;

OY 1 MWQIKPTNEKFTFLTAAGKLAIVQFSSKRCGCKRMFPVHELAE-----TCH 50
Db 77 VTEVKNKT--FWPIYKAAGDKPVYLDMTQWCGPCKVIAPLYEELSLQYLDVFLKLDGN 134

OY 51 -----IKTPTFQMFKKSOKV 66
Db 135 QDNKSLAKELGITKVPFTFKLDKNIV 161

RESULT 12
S04661
thioedoxin f precursor - spinach
C:Species: Spinacia oleracea (spinach)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 11-Jun-1999
C:Accession: S04661; A20273
R:Kamo, M.; Tsugita, A.; Miessner, C.; Wedel, N.; Bartling, D.; Herrmann, R.G.; Aguilar,
Eur. J. Biochem. 182, 315-322, 1989
A:Title: Primary structure of spinach-chloroplast thioedoxin f. Protein sequencing and
A:Reference number: S04661; MUID:89289715
A:Accession: S04661
A:Molecule type: mRNA
A:Residues: 1-190 <KAM>
A:Cross-references: EMBL:X14959; NID:g21343; PIDN:CAA33082.1; PID:g21344
A:Note: part of this sequence was confirmed by protein sequencing
R:Tsugita, A.; Maeda, K.; Schnermann, P.
Biochem. Biophys. Res. Commun. 115, 1-7, 1983
A:Title: Spinach chloroplast thioedoxin in evolutionary drift.
A:Reference number: A20273; MUID:83308759
A:Accession: A20273
A:Molecule type: protein
A:Residues: 106-119, 'NGDKKATQHLGVQDA', 120 <TSU>
A:Note: this sequence has been revised in reference S04661
C:Comment: Thioedoxin f activates fructose 1,6-bisphosphatase during photosynthesis.
C:Superfamily: thioedoxin; thioedoxin homology
C:Keywords: chloroplast; redox-active disulfide
F:1-77/Domain: transit peptide (chloroplast) #status predicted <TNP>
F:78-190/Product: thioedoxin f #status predicted <MAT>
F:91-115/Domain: thioedoxin homology <THR>
F:114-117/Disulfide bonds: redox-active #status predicted

```

```

Query Match 18.4%; Score 103.5; DB 2; Length 190;
Best Local Similarity 29.9%; Pred. No. 0.00063;
Matches 26; Conservative 16; Mismatches 22; Indels 23; Gaps 3;

OY 1 MWQIKPTNEKFTFLTAAGKLAIVQFSSKRCGCKRMFPVHELAE-----TCH 50
Db 85 VTEVKNKT--FWPIYKAAGDKPVYLDMTQWCGPCKVIAPLYEELSLQYLDVFLKLDGN 142

OY 51 -----IKTPTFQMFKKSOKV 66
Db 143 QENKTLAKELGIRVVPFTFKLENSV 169

RESULT 13
B96721
probable thioedoxin T17F3.9 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B96721
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Aloni
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzic
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tall
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: B96721
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-148 <STO>
A:Cross-references: GB:AE005173; NID:g6358805; PIDN:AAF07385.1; GSPDB:GN00141
C:Genetics:
A:Gene: T17F3.9
A:Map position: 1

Query Match 18.3%; Score 103; DB 2; Length 148;
Best Local Similarity 27.5%; Pred. No. 0.00057;
Matches 25; Conservative 16; Mismatches 26; Indels 22; Gaps 2;

OY 5 IKDTNEKFTFLTAAGKLAIVQFSSKRCGCKRMFPVHELAE-----TCH 47
Db 42 IKNNQWKSRLNALKDTNKLIVIEFTAKWCGPCKTLEFLAKYTVDFVKIDVVL 101

OY 48 -----TCHIKITPTFQMFKKSOKVTLFSRIK 73
Db 102 MSVMNEFNLSTLPAIVKRGREVDMMVGK 132

RESULT 14
TXBY1
thioedoxin I - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein G7746; protein YGR209C; thioedoxin 2
C:Species: Saccharomyces cerevisiae
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jun-2000
C:Accession: S15049; B39847; S05793; S53932; S61947; S64531; S63858; A38669
R:Gan, Z.R.
J. Biol. Chem. 266, 1692-1696, 1991
A:Title: Yeast thioedoxin genes.
A:Reference number: A38669; MUID:91107668
A:Accession: S15049
A:Molecule type: DNA
A:Residues: 1-104 <GAN>
A:Cross-references: GB:M59168; NID:g173025; PIDN:AAA35170.1; PID:g173026
R:Miller, E.G.D.
J. Biol. Chem. 266, 9194-9202, 1991
A:Title: Thioedoxin deficiency in yeast prolongs S phase and shortens the G1 interval

```

A:Reference number: A39847; MUID:91225027
A:Accession: B39847
A:Molecule type: DNA
A:Residues: 1-104 <MDL>
A:Cross-references: GB:M62648; NID:9173049; PIDN:AAA5178.1; PID:9173050
R:Hall, D.E.; Balessten, A.; Holmgren, A.; Reichard, P.
Eur. J. Biochem. 23, 328-335, 1971
A:Title: Yeast thioredoxin. Amino acid sequence around the active-center disulfide of th
A:Reference number: S05793; MUID:72100583
A:Accession: S05793
A:Molecule type: protein
A:Residues: 2:27-43;98-104 <HAL>
A:Note: the sequence from the summary and from Fig. 5 is inconsistent with that from pag
R:Guerrero, P.; Barreiros, T.; Soares, H.; Cyrne, L.; Maia e Silva, A.; Rodrigues-Pousa
submitted to the EMBL Data Library, April 1995
A:Description: Sequencing of a 17.6 kb segment on the right arm of yeast chromosome VII
es, of the human.
A:Reference number: S53922
A:Accession: S53922
A:Molecule type: DNA
A:Residues: 1-104 <GDE>
A:Cross-references: EMBL:Z49133; NID:9790489; PIDN:CAA69002.1; PID:9790500
A:Experimental source: strain S286C
R:Song, J.M.; Cheung, E.; Rabinowitz, J.C.
submitted to the EMBL Data Library, November 1995
A:Description: Analysis of the 15.6-kb fragment encompassing the *ADP3* gene.
A:Reference number: S61947
A:Accession: S61947
A:Molecule type: DNA
A:Residues: 1-104 <SON>
A:Cross-references: EMBL:U040843; NID:91165213; PIDN:AAA85584.1; PID:91165214
A:Experimental source: Strain GPF88
R:Guerrero, P.; Barreiros, T.; Cyrne, L.; Soares, H.; Maia e Silva, A.; Rodrigues-Pousa
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64517
A:Accession: S64517
A:Molecule type: DNA
A:Residues: 1-104 <GWP>
A:Cross-references: EMBL:Z72994; NID:91333374; PIDN:CAA97236.1; PID:91323375; GSPDB:GNOC
A:Experimental source: strain S286C
R:Guerrero, P.; Barreiros, T.; Soares, H.; Cyrne, L.; Maia e Silva, A.; Rodrigues-Pousa
Yeast 12, 273-280, 1996
A:Title: Sequencing of a 17.6 kb segment on the right arm of yeast chromosome VII reveal
terial electron-transferring flavoproteins (beta-chain) and of the *Escherichia coli* phos
A:Reference number: S63848; MUID:97060019
A:Accession: S63848
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-104 <GUP>
A:Cross-references: EMBL:Z49133; NID:9790489; PIDN:CAA69002.1; PID:9790500
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, April 1995
C:Genetics:
A:Gene: SGD:TRX2; TR-1; MIPS:YGR209C
A:Cross-references: SGD:S0003441; MIPS:YGR209C
A:Map position: 7R
C:Superfamily: thioredoxin; thioredoxin homology
C:Keywords: redox-active disulfide
F:2-104/Product: thioredoxin I #status experimental <MAT>
F:9-91/Domain: thioredoxin homology <THR>
F:31-34/Disulfide bonds: redox-active #status experimental

Query Match 18.2%; Score 102.5; DB 1; Length 104;
Best Local Similarity 28.7%; Pred. No. 0.00047;
Matches 25; Conservative 15; Mismatches 26; Indels 21; Gaps 2;

OY 1 MWIITKDNNEKFTLTAGHKLAVVQFSKRCGPKRMFPVHFLAE----- 47
DB 1 MVTOLKASSETDSAL-ASGDKLVVDFPRTWCGPCKMTAPMEIKFAEQYSDAATFKLDVD 59
OY 48 -----TCHIKTIPTFQMFRRKSQKVT 67
DB 60 EVSDVAQKAQVSSMPTLIFYGKGKEVT 86

RESULT 15
S47867
thioredoxin-like protein - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S47867
R:Salz, H.R.; Flickinger, T.W.; Miltendorf, E.; Pelliscena-Palle, A.; Petschek, J.P.;
Genetics 136, 1075-1086, 1994
A:Title: The *Drosophila* maternal effect locus *deadhead* encodes a thioredoxin homolog
A:Reference number: S47867; MUID:94274010
A:Accession: S47867
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-107 <SAL>
A:Cross-references: EMBL:L27072; NID:9435591; PIDN:AAA28937.1; PID:9435963
C:Genetics:
A:Gene: FlyBase:dhd
A:Cross-references: FlyBase:FBgn0011761
C:Superfamily: thioredoxin; thioredoxin homology
C:Keywords: redox-active disulfide
F:8-92/Domain: thioredoxin homology <THR>
F:31-34/Disulfide bonds: redox-active #status predicted

Query Match 18.1%; Score 101.5; DB 1; Length 107;
Best Local Similarity 22.4%; Pred. No. 0.00062;
Matches 26; Conservative 17; Mismatches 30; Indels 43; Gaps 2;

OY 5 IKDNEKFTLTAGHKLAVVQFSKRCGPKRM-----FPVPH 43
DB 4 VRTMNDYHKRIEADDDKLVDFYATWCGPCKEMESTYKSLARRYSSKAVVLIKIDVDKFE 63
OY 44 ELAETCHIKTIPTFQMFRRKSQKVTLSRIKRIICCYRSGFMSNLIFFCGADAKKL 99
DB 64 ELTERYKVRSMPTFVFLKQNRRLA-----SPAGADEKRL 97

Search completed: June 8, 2002, 04:29:51
Job time: 3835 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2002, 03:21:11 ; Search time 32.08 Seconds
(without alignments)
81.469 Million cell updates/sec

Title: US-09-746-783-19

Perfect score: 562

Sequence: 1 MWQIKDKNEKFTLLAKGH.....IFEFGADAKKLEAKTOELM 107

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	190	33.8	210	4	US-08-180-371-6
2	180	33.8	210	5	PCT-US92-05707-6
3	188	33.5	105	2	US-08-826-910-3
4	185	32.9	104	6	5210073-1
5	185	32.9	105	2	US-08-826-910-4
6	118	21.0	130	2	US-08-826-910-1
7	85.5	15.2	106	1	US-08-386-729A-10
8	84.5	15.0	111	4	US-09-264-419C-6
9	77	13.7	1285	4	US-08-974-949A-600
10	75	13.3	521	2	US-08-557-122A-32
11	75	13.3	521	4	US-09-262-666-32
12	74	13.2	129	4	US-08-557-122A-11
13	74	13.2	129	4	US-09-262-666-11
14	74	13.2	177	4	US-09-272-342B-6
15	74	13.2	485	1	US-08-068-395A-1
16	74	13.2	485	1	US-08-464-365-1
17	74	13.2	505	1	US-08-068-395A-3
18	74	13.2	505	1	US-08-464-365-3
19	73	13.0	131	2	US-08-557-122A-10
20	73	13.0	131	4	US-09-262-666-10
21	73	13.0	141	2	US-08-557-122A-9
22	73	13.0	141	4	US-09-262-666-9
23	73	13.0	143	2	US-08-557-122A-8
24	73	13.0	143	4	US-09-262-666-8
25	73	13.0	163	4	US-08-557-122A-7
26	73	13.0	163	4	US-09-262-666-7
27	73	13.0	174	2	US-08-557-122A-6

28	73	13.0	174	4	US-09-262-666-6	Sequence 6, Appl1
29	73	13.0	200	2	US-08-557-122A-12	Sequence 12, Appl
30	73	13.0	200	4	US-09-262-666-12	Sequence 12, Appl
31	73	13.0	281	2	US-08-557-122A-5	Sequence 5, Appl1
32	73	13.0	281	4	US-09-262-666-5	Sequence 5, Appl1
33	73	13.0	511	2	US-08-557-122A-4	Sequence 4, Appl1
34	73	13.0	511	4	US-09-262-666-4	Sequence 4, Appl1
35	73	13.0	515	2	US-08-557-122A-3	Sequence 3, Appl1
36	73	13.0	515	2	US-08-557-122A-34	Sequence 34, Appl
37	73	13.0	515	4	US-09-262-666-3	Sequence 3, Appl1
38	73	13.0	515	4	US-09-262-666-34	Sequence 34, Appl
39	73	13.0	530	2	US-08-557-122A-35	Sequence 35, Appl
40	73	13.0	530	4	US-09-262-666-35	Sequence 35, Appl
41	72.5	12.9	109	1	US-07-745-382-22	Sequence 22, Appl
42	72.5	12.9	109	1	US-07-921-848-22	Sequence 22, Appl
43	72.5	12.9	109	1	US-08-165-301A-22	Sequence 22, Appl
44	72.5	12.9	109	4	US-08-810-436-22	Sequence 22, Appl
45	72.5	12.9	109	4	US-09-485-885-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-08-180-371-6
; Sequence 6, Application US/08180371
; Patent No. 6254861
; GENERAL INFORMATION:
; APPLICANT: Choudhury, Chandra
; TITLE OF INVENTION: Hematopoietic Growth Factor Derived
; TITLE OF INVENTION: from T Lymphocytes and Methods of Use Therefor
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/180,371
; FILING DATE: 12-JAN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 906866
; FILING DATE: 01 July 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gruber, Lewis S.
; REGISTRATION NUMBER: 30,060
; REFERENCE/DOCKET NUMBER: 27620/31668
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 210 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-180-371-6

Query Match 33.8%; Score 190; DB 4; Length 210;
Best Local Similarity 36.2%; Pred. No. 1.5e-17;
Matches 46; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

CLONE:
US-08-826-910-4

1070-90-00

```
QY      10 EFKTFLTAAGHKLAVNQFSSKRCGCPCKRMPEVFHEL-----AETC 49
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      12 DFOPELSGAGSRILAVKFTTMRGCGPCRLRIAPAFSSMSKNKYPQAVFLEVDVHOCGGTAATN 71
```



```

1      PRIOR APPLICATION DATA:
2      APPLICATION NUMBER: US 08/724,643
3      FILING DATE: 01-OCT-1996
4      PRIOR APPLICATION DATA:
5      APPLICATION NUMBER: US 08/844,419
6      FILING DATE: 18-APR-1997
7      PRIOR APPLICATION DATA:
8      APPLICATION NUMBER: US 08/846,017
9      FILING DATE: 25-APR-1997
10     PRIOR APPLICATION DATA:
11     APPLICATION NUMBER: US 08/851,843
12     FILING DATE: 06-MAY-1997
13     PRIOR APPLICATION DATA:
14     APPLICATION NUMBER: US 08/854,050
15     FILING DATE: 09-MAY-1997
16     PRIOR APPLICATION DATA:
17     APPLICATION NUMBER: US 08/911,312
18     FILING DATE: 14-AUG-1997
19     PRIOR APPLICATION DATA:
20     APPLICATION NUMBER: US 08/912,951
21     FILING DATE: 14-AUG-1997
22     PRIOR APPLICATION DATA:
23     APPLICATION NUMBER: US 08/915,503
24     FILING DATE: 14-AUG-1997
25     PRIOR APPLICATION DATA:
26     APPLICATION NUMBER: WO PCT/US97/17618
27     FILING DATE: 01-OCT-1997
28     PRIOR APPLICATION DATA:
29     APPLICATION NUMBER: WO PCT/US97/17885
30     FILING DATE: 01-OCT-1997
31     ATTORNEY/AGENT INFORMATION:
32     NAME: Apple, Randolph Ted
33     REGISTRATION NUMBER: 36,429
34     REFERENCE/DOCKET NUMBER: 015389-002610US
35     TELECOMMUNICATION INFORMATION:
36     TELEPHONE: (415) 576-0200
37     TELEFAX: (415) 576-0300
38     INFORMATION FOR SEQ ID NO: 600:
39     SEQUENCE CHARACTERISTICS:
40     LENGTH: 1285 amino acids
41     TYPE: amino acid
42     STRANDEDNESS:
43     TOPOLOGY: linear
44     MOLECULE TYPE: protein
45     FEATURE:
46     NAME/KEY: Protein
47     LOCATION: 1..1285
48     OTHER INFORMATION:
49     OTHER INFORMATION:
50     OTHER INFORMATION:
51     /note="fusion protein composed of
52     enterokinase cleavable, His tagged
53     thiodoxin moiety and full length hTTR"
54     US-08-974-5494-600

```

```

Query Match          13.7%; Score 77; DB 4; Length 1285;
Best Local Similarity 23.9%; Pred. No. 0.2;
Matches 27; Conservative 15; Mismatches 33; Indels 38; Gaps 4.

QY      24 VVQESSKRCGCKRMFPPEHLEATCH-----IKITPFQMEFK 62
      : 1 : 11111 : 1 : 11 :
Db      25 LVDFPAHHCQCKMLAPFLDELADYQCKLYAKLRIDHNPGTARKYGRIGIPTLLFFKN 84
QY      63 SQ----KYLFSR--IKRIICYRSGFMSN-----LIFEECGADARK 98
      : 11 : 11 : 11 : 1111 : 1 :
Db      85 GEVAATKVGALSKGQLKKEFLDANLAGSGSDDDKVPMPHELEIFEPAAASTQR 137

RESULT 10
US-08-557-122A-32
; Sequence 32, Application US/08557122A
; Patent No. 5879664
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Mailand
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase

```

NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58796640 No. 5879664disk of No. 5879664th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,122A
FILING DATE: 11-Dec-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3980.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-557-122A-32

```

Query Match Similarity 13.3%; Score 75; DB 2; Length 521;
Best Local Similarity 25.6%; Pred. No. 0.11;
Matches 20; Conservative 12; Mismatches 22; Indels 24; Gaps 2

QY      8  TNEFTFLTAACHKLAIVQFSSKRCGCPCKRMPVPEHELAETCHIKTI----- 54
      1 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      39  TDSFNEIYDS--HDLVLAIEFFAPQWCHCKNMAPEYKAAETLVENKNTLLAQIDCTENQDL 96
      1 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      55  -----PTFQMRKKS 63
      1 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      97  CMEHNIPGFPSLKIIFKNS 114

RESULT 11
US-09-262-666-32
; Sequence 32, Application US/09262666
; Patent No. 6346244
; GENERAL INFORMATION:
; APPLICANT: HJort, Carsten Mølland
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 63462440 No. 6346244disk of No. 6346244th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/262,666
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,122

```

;; GENERAL INFORMATION:
; APPLICANT: YU, YEON-GYU

AFFILIATION NUMBER: JP 3-44014

Search completed: June 8, 2002, 04:28:49
Job time: 4058 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2002, 04:28:51 ; Search time 89.36 Seconds
(without alignments)
207.145 Million cell updates/sec

Title: US-09-746-783-19

Perfect score: 562

Sequence: 1 MVQIKDNEFKTFLTAGH.....IPEFGADAKKLEAKTQELM 107

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	313	55.7	127	11	09CQ96 mus musculus
2	202	35.9	84	4	060744 mus sapien
3	188	33.5	105	4	096K13 mus sapien
4	187	33.3	105	6	095JF9 sus scrofa
5	185	32.9	105	6	097508 equus caball
6	185	32.9	105	6	09BDJ3 gallitrix
7	182	32.4	105	11	09DBR0 mus musculus
8	179	31.9	105	13	098TX1 mus musculus
9	165	29.4	112	4	09H0L8 mus sapien
10	160	28.5	486	4	096RX3 mus sapien
11	146	26.0	107	13	09DGI3 mus sapien
12	135.5	24.1	106	5	0963B4 mus sapien
13	134	23.8	287	5	09VRP3 aedes aegypt
14	133.5	23.8	107	5	09NGZ1 anopheles g
15	132.5	23.6	104	5	0962B7 branchiosto
16	124	22.1	284	5	09N2K6 caenorhabdit

17	123.5	22.0	106	5	09U515 manduca sex
18	118	21.0	289	11	070379 mus musculus
19	118	21.0	289	11	0920J4 rattus norv
20	117.5	20.9	106	5	09V429 drosophila
21	115.5	20.6	104	5	09N1R2 plasmodium
22	114	20.3	106	5	096952 geodia cydo
23	110.5	19.7	104	5	09NFK9 plasmodium
24	105.5	18.8	114	5	0955W4 drosophila
25	103	18.3	148	10	09CAS1 arabisopsis
26	102.5	18.2	101	10	022031 cyanidium c
27	102.5	18.2	106	5	0908R3 schistosoma
28	99.5	17.7	245	10	09XFI1 arabisopsis
29	98.5	17.5	178	10	09M8R5 arabisopsis
30	97	17.3	385	10	093X83 vitis labru
31	96	17.1	140	17	096YQ0 sulfolobus
32	95.5	17.0	121	3	094504 schizosacch
33	94	16.7	131	10	09S753 phalaris co
34	94	16.7	131	10	09SMG6 hordeum bul
35	94	16.7	131	10	09SMG4 lolium pere
36	93.5	16.6	98	5	025549 naegleria f
37	93.5	16.6	275	10	064554 arabisopsis
38	91.5	16.3	304	5	09N357 caenorhabdit
39	91	16.2	215	10	09XFI10 arabisopsis
40	91	16.2	221	10	09S2I7 arabisopsis
41	89	15.8	125	10	09LIX4 triticum ae
42	89	15.8	130	10	064395 triticum tu
43	89	15.8	140	10	09C9Y6 arabisopsis
44	88.5	15.7	107	5	09TX18 caenorhabdit
45	88.5	15.7	111	5	09W4D6 drosophila

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	127 AA.
09CQ96	09CQ96			
AC	09CQ96:			
DT	01-JUN-2001 (TReMBLrel. 17, Created)			
DT	01-JUN-2001 (TReMBLrel. 17, Last sequence update)			
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)			
DE	4930429J24RIK PROTEIN.			
GN	4930429J24RIK.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=TESTIS;			
RC	MEDLINE=21085660; PubMed=11217851;			
RA	Atakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staudili F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Homberters P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,			
RA	Hayashizaki Y.;			
RT	"Functional annotation of a full-length mouse cDNA collection."			
RL	Nature 409:685-690(2001).			
DR	EMBL; AK015240; BAB29760.1; -			
DR	EMBL; AK006683; BAB24702.1; -			

RESULT	5		
097508			
ID	097508	PRELIMINARY;	PRT; 105 AA.
AC	097508;		
DT	01-MAY-1999 (TREMBLrel. 10, Created)		
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	THIOREDOXIN.		
CN	TXN.		
OS	Equus caballus (Horse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Perissodactyla; Equidae; Equus.		
OX	NCBI_TaxId=9796;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=THOROUGHRED;		
RA	Tajima Y., Ishida N.;		
RT	"Molecular Cloning of Equine Thioredoxin.";		
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB022431; BAA37154.1; -.		
DR	HSSP; P10599; IERV.		
DR	InterPro; IPR000063; Thlored.		
DR	Pfam; PF00085; thlored. 1.		
DR	PRINTS; PR00421; THIOREDOXIN.		
DR	PROSITE; PS00194; THIOREDOXIN; 1.		
DR	Redox-active center.		
QC	SEQUENCE 105 AA; 11736 MW; 04F9E35DA8FB5C4 CRC64;		

[illegible]

DR PROSITE; PS00194; THIOREDOXIN; 1.
KM Redox-active center.
SQ SEQUENCE 105 AA; 11757 MW; 649AA18C9CB81B38 CRC64;

	Query Match	Similarity	32.9%	Score	185:	DB	6:	Length	105:	
	Best Local	Similarity	35.4%	Pred.	No.	3e-15:				
	Matches	45:	Conservative	11:	Mismatches	29:	Indels	42:	Gaps	27:
OY	1	VWQIIKIDTNEFKETLLAAGHKLAVVQFSSRCGCPCKRMFPVHELAE-----	47							
Dd	1	WKVKIDSKDAPFOEALDAGDKLVVVDFATMCGPKMIKPFHSLSEKSYNVFLVEDVD	60							
OY	48	-TCHIKITTFQMFKKSCQVTLFSRKRIICCRSGFMNLTRFCGDAAKLE	100							
Dd	61	DCDDVASCEKCMPTPEFFPKGRKV-----EFSGAKKEKLE	98							
OY	101	AKTQELM	107							
Dd	99	ATINEFV	105							
RESULT	7									
ID	O9D8R0	PRELIMINARY;	PRT:	105 AA.						
AC	O9D8R0:									
DT	01-JUN-2001	(TREMBLrel. 17, Created)								
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)								
DT	01-DEC-2001	(TREMBLrel. 19, last annotation update)								
DE	10 DAY OLD MALE PANCREAS CDNA, RIKEN FULL-LENGTH ENRICHTED LIBRARY,									
DE	CLONE:1810045N15, FULL INSERT SEQUENCE.									
GN	TXN.									
OS	Mus musculus (Mouse).									
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;									
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.									
OX	NCBI_TaxID=10090;									
RN	[1]									
RP	SEQUENCE FROM N.A.									
RC	STRAIN=C57BL/6J; TISSUE=PANCREAS;									
RX	MEDLINE=21085660; PubMed=11217851;									
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,									
RA	Atikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,									
RA	Alkawa K., Iwama M., Nishi K., Kiyosawa H., Kondo S., Yamanaa I.,									
RA	Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,									
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,									
RA	Fischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,									
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,									
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,									
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,									
RA	Blake J., Boiffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,									
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,									
RA	Gustincich S., Hill D., Hofman M., Hume D.A., Kamuya M., Lee N.H.,									
RA	Lions P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,									
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,									
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,									
RA	Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,									
RA	Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,									
RA	Hayashizaki Y.;									
RL	"Functional annotation of a full-length mouse cDNA collection.":									
RL	Nature 409:665-690(2001).									
DR	EMBL; AK007790; BABS2556.1; -									
DR	HSSP; P10599; IERT.									
DR	MGI; MGI:98874; Txn.									
DR	InterPro; IPR000063; Thioded.									
DR	pfam; PF00085; Thioded. 1.									
DR	PRINTS; PR00421; THIOREDOXIN.									
DR	PROSITE; PS00194; THIOREDOXIN; 1.									
SW	Redox-active center.									
SW	SEQUENCE 105 AA; 11691 MW; 3A2B9535935E252DA CRC64;									

Query Match	32.4%;	Score 182;	DB 11;	Length 105;
Best Local Similarity	35.5%;	Pred. No. 7.1e-15;		

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Miranda-Vizuet A.;
 RT "Drosophila melanogaster homolog of human txl gene,"
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003565; AAF50750.1; -
 DR EMBL; AF143404; AAF66635.1; -
 DR HSSP; O43396; 1GH2.
 DR FLYBase; FBgn0035631; Txl.
 DR InterPro; IPR000362; Fumarate_lyase.
 DR InterPro; IPR000063; Thioired.
 DR Pfam; PF00085; Thioired.1.
 DR PRINTS; PR00421; THIOREDOXIN.
 DR PROSITE; PS00163; FUMARATE_LYASES; UNKNOWN_1.
 DR PROSITE; PS00194; THIOREDOXIN; 1.
 KW Redox-active center.
 FT CONFLICT 149 150 LA -> WP (IN REF. 2).
 FT CONFLICT 177 177 H -> Q (IN REF. 2).
 SQ SEQUENCE 287 AA; 31728 MW; BB316D91A9432A6 CRC64;

Query Match 23.8%; Score 134; DB 5; Length 287;
 Best Local Similarity 27.4%; Pred. No. 1.8e-08;
 Matches 34; Conservative 17; Mismatches 31; Indels 42; Gaps 4;

QY 2 VQIITDNEFKTFLTAGHKLAVVQFSSKRCGPCRMFPVHLEA-----AET 48
 Db 3 VRVINDESHFOALAOAOLVVDFTASWCGPCRMIPETFTPKYKPAFLKVDYDK 62
 QY 49 CH-----IKTPTFQMKRSQKVTLSFKIRIKICYSRSGFMSNLIFECGADAKKLEA 101
 Db 63 QDPAAGCGVAMPTFIYRNRTKI---DRVQ-----GADVNGLEA 100
 QY 102 KTOE 105
 Db 101 KIOE 104

RESULT 14
 Q9NGZ1 PRELIMINARY; PRT; 107 AA.
 AC Q9NGZ1;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE THIOREDOXIN 1.
 OS Anopheles gambiae (African malaria mosquito).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 OC Anopheles.
 OX NCBI_Taxid=7165;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kanok S., Schluter R.H., Becker K.;
 RT "Thioiredoxin 1 of the malaria vector *Anopheles gambiae*,"
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF236124; AAF68382.1; -
 DR HSSP; P10599; IERV.
 DR InterPro; IPR000063; Thioired.
 DR Pfam; PF00085; Thioired.1.
 DR PRINTS; PR00421; THIOREDOXIN.
 DR PROSITE; PS00194; THIOREDOXIN; 1.
 KW Redox-active center.
 SQ SEQUENCE 107 AA; 12085 MW; 8B8E404B18F9A0E7 CRC64;

Query Match 23.8%; Score 133.5; DB 5; Length 107;
 Best Local Similarity 27.8%; Pred. No. 7.3e-09;
 Matches 35; Conservative 18; Mismatches 30; Indels 43; Gaps 2;

QY 1 MVQIITDNEFKTFLTAGHKLAVVQFSSKRCGPCRMFPVHLEA-----AET 42
 Db 1 MVYMKDSEDFNNKLEAGDGLVVDFTATWCGPCRMIPETFTPKYKPAFLKVDYDK 60
 QY 43 ---HELAETCHIKTPTFQMKRSQKVTLSFKIRIKICYSRSGFMSNLIFECGADAKK 99
 Db 61 DECELAQVYNASMPTELFIRKE-----VVGQFSGANAEKL 98
 QY 100 EAKTOE 105
 Db 99 ENFTQO 104

RESULT 15
 Q962B7 PRELIMINARY; PRT; 104 AA.
 AC Q962B7;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE THIOREDOXIN.
 OS Branchiostoma belcheri (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 OX NCBI_Taxid=7741;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Huang X., Wang X., Li L., Han H., Yang H., Zhang H.;
 RT "Cloning of thioiredoxin gene in amphioxus,"
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY037864; AAK72483.1; -
 SQ SEQUENCE 104 AA; 11431 MW; 07D453D14FC68A41 CRC64;

Query Match 23.6%; Score 132.5; DB 5; Length 104;
 Best Local Similarity 31.1%; Pred. No. 9.5e-09;
 Matches 37; Conservative 11; Mismatches 26; Indels 43; Gaps 4;

QY 1 MVQIITDNEFKTFLTAGHKLAVVQFSSKRCGPCRMFPVHLEA-----AET 47
 Db 1 MVQMETKAAPDKLL-AETDKLIYVDFTASWCGPCRMIPETFTPKYKPAFLKVDYDK 59
 QY 48 -----TCHIKTPTFQMKRSQKVTLSFKIRIKICYSRSGFMSNLIFECGADAKK 99
 Db 60 ANDETAGAGISAMPTFH-----CYRNGAK---VEELGASEEKL 96

Search completed: June 8, 2002, 04:36:57
 Job time: 486 sec

Sat Jun 8 10:14:14 2002

us-09-746-783-19.rsp
